Unravelling taxonomic uncertainties among balloon vine species within the genus *Cardiospermum* using a molecular approach

by

Enelge Gildenhuys

Thesis presented in fulfilment of the requirements for the

degree Masters of Science in Botany at Stellenbosch University

Supervisor: Dr. Johannes J. Le Roux

Co-supervisor: Dr. Allan G. Ellis and Dr. Scott P. Carroll

December 2013

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2013

Thesis Abstract

Alien invasive species are a global concern not only threating biodiversity, but also negatively impacting regional economies. Consequently research aimed at understanding the invasion process is crucial for management of invasive alien species to reduce these impacts. Identifying accurate natal ranges is the key first step for efficient alien plant management, especially biological control. Such knowledge may assist in selecting host-specific biological control agents and help prevent non-target impact on native species. A case in point is the genus *Cardiospermum* of which species have been extensively moved around the globe with two species now being possible widespread invaders (*C. grandiflorum* and *C. halicacabum*), and a third species (*C. corindum*) having significant invasion potential. However, in some regions the native statuses of these species are not clear, hampering management. In order to minimize potential non-target impacts it is prudent to determine the relationships between *Cardiospermum* taxa in unknown native ranges prior to the release of biological control agents.

With this thesis I aim to review available literature on the genus *Cardiospermum*, determine potential spread to suitable habitats globally for selected taxa, resolve unknown native ranges in southern Africa, and investigate rapid speciation of an endemic Namib Desert species of balloon vine.

Through exhaustive searches for available literature on the genus *Cardiospermum* I report on the biology and ecology of selected species within the genus, with special focus on the most widespread species. Specifically, using species distribution modelling I investigate the potential spread of *C. halicacabum*, *C. corindum* and *C. grandiflorum* globally and assess the accuracy with which this can be determined using known native ranges to predict current

alien ranges. Results indicate that these species have significant potential to spread, though highlighting that species distribution modelling over-fit predicted suitable ranges when using native range data alone.

To resolve uncertain native ranges of selected species within the genus I reconstructed a dated multi-gene phylogeny and chloroplast haplotype network to investigate relationships within and among *Cardiospermum* species. I conclude that *C. halicacabum* is non-native in southern Africa due to polyphyletic relationships among accessions, while *C. corindum* is native due to natural long distance dispersal as indicated by a single monophyletic southern African clade.

The close relationship observed between *C. corindum* and the African species, *C. pechuelii*, lead to a population-level genetic investigation of these two taxa. For this, I constructed a multi-gene phylogeny, chloroplast haplotype network and conducted population genetic diversity analyses which all indicated that the African arid adapted taxa, *C. pechuelii*, endemic to the Namib Desert of Namibia, evolved from *C. corindum* following long distance dispersal from South America.

This study firstly provides insights into the ecology and biology of *Cardiospermum*, warning against further introduction due to potential spread. Secondly I resolved the unknown native statutes of *Cardiospermum* species in southern Africa and, lastly, illustrated the origin of African *C. pechuelii*. My research findings have major implications for on-going biological research against *C. grandiflorum* in South Africa and cautions strongly against the release of already-identified biological control agents due to their potential non-target impacts.

Tesis Opsomming

Indringerspesies is 'n wêreldwye probleem as gevolg van hul direkte bedreiging van biodiversiteit asook hul negatiewe impakte op streeksekonomieë. Gevolglik is navorsing aangaande die indringingsproses van uiterste belang om doeltreffende beheer van indringerspesies te verseker. Die bepaling van die inheemse herkoms van indringerpopulasies is die eerste belangrike stap tot doeltreffende beheer, veral biologiese beheer. Hierdie informasie kan help met die identifisering van gasheer-spesifieke biologiese agente en om potensiële nie-teiken impakte op inheemse plant spesies verhoed. ‘n Voorbeeld is ballon rankplante in die genus *Cardiospermum*, waarvan verskeie spesies wêreldwyd verspreid is, met die gevolg dat minstens twee spesies (*C. halicacabum* en *C. grandiflorum*) nou moontlike wydverspreide indringers is en ‘n derde spesie (*C. corindum*) ook potensiaal toon as ‘n indringerspesie**.** In sommige streke is die in- of uitheemse statusse van hierdie spesies onbekend en belemmer gevolglik hul doeltreffende beheer.

Met hierdie tesis streef ek om die genus *Cardiospermum* te hersien en potensiële verspreiding van geselekteerde spesies na geskikte habitatte te bepaal, om onbekende inheemse streke binne Suider-Afrika op te los en laastens, om die herkoms van ‘n ware Afrika spesie (*C. pechuelii*) te ondersoek.

Na deeglike ondersoek van beskikbare literatuur doen ek verslag omtrent die biologie en ekologie van spesies binne *Cardiospermum*, met spesiale fokus op die mees wydverspreidste spesies. Deur gebruik te maak van spesies verspreidings modelle bestudeer ek die potensiaal van verdere verspreiding wêreldwyd vir *C. halicacabum*, *C. corindum* en *C. grandiflorum* en rapporteer ook omtrent die akkuraatheid van hierdie tegniek. Resultate toon

dat alle spesies aansienlike potensiaal toon om verder te versprei, maar beklemtoon ook dat hierdie metode inheemse streek data “oor-aanpas”.

Om enige onsekerhede aangaande inheemse streke op te los, het ek ‘n gedateerde multi-geen filogenie en ‘n chloroplas haplotipe netwerk opgetrek om die verhoudings binne en tussen geselekteerde *Cardiospermum* spesies te bepaal. Ek kom tot gevolgtrekking dat *C. halicacabum* nie inheems in Suider-Afrika is nie, as gevolg van polifiletiese verhoudings, en dat *C. corindum* inheems is en moontlik ook ‘n voorbeeld van natuurlike langafstand verspreiding, soos aangedui deur ‘n enkele monofiletiese groep.

Die naverwante filogetiese verhouding tussen *C. corindum* en *C. pechuelii* het daartoe gelei dat ek die herkoms van *C. pechuelii* in Afrika ondersoek het. ‘n Multi-geen filogenie, chloroplas haplotipe netwerk en populasie genetiese diversiteit analises het aangedui dat *C. pechuelii* endemies is tot die Namib Woestyn van Namibië, en die produk is van lang afstand verspreiding van *C. corindum* vanaf Suid-Amerika.

Hierdie studie gee eerstens insig tot die ekologie en biologie van *Cardiospermum* en waarsku teen verdere verspreiding. Tweedens dui my studie ook die opgeloste inheemse streke in Suider-Afrika aan van sekere *Cardiospermum* spesies, en laastens, illustreer ek die herkoms van *C. pehuelii*. My navorsing het baie belangrike gevolge vir die huidige biologiese beheer program teen *C. grandiflorum* in Suid-Afrika en waarsku teen die vrystelling van reeds geïndentifiseerde biologiese beheer agente weens potensiële nie-teiken impakte op inheemse spesies.

Acknowledgements

First and most important I want to thank my supervisor, Dr. Jaco Le Roux, for his supervision, guidance and patience throughout the duration of this project. I also thank my co-supervisors, Dr. Allan Ellis and Dr. Scott Carroll, for their helpful insights and feedback on the project.

I thank everyone who helped with sampling, especially David Simelane and Andries Fourie who sampled *Cardiospermum* in South America. I also thank Hannes Wiese and Mandelise van Loggerenberg for field assistance, ensuring that both sampling trips were extremely memorable.

I acknowledge the financial support of the DST-NRF Centre of Excellence for Invasion Biology and the NRF Scarce Skills scholarship programme for funding this project.

Lastly I thank my mother, farther, sister, brother-in-law, all my friends and especially Chris de Jongh for their words of encouragement throughout my Masters, without them it would've been a lot harder if not impossible.

Contents

List of Figures

Chapter 1

Chapter 2

introductions from different sources in the native range, B: Rare long distance dispersal would lead to a single group of closely related African haplotypes along with a phylogeny illustrating a monophyletic clade of southern African samples in an overall dominant South American clade and C: Gondwanan distribution would lead to highly structured and diverged haplotypes for South American and southern African samples and a phylogeny with a deep node split and geographic monophyly since time of divergence, ca. 90 MYA, ample time for unique genetic mutations to accumulate.

Figure 2: *rpl32* haplotype network of native, invasive and unknown range *Cardiospermum* species. Different countries are represented by different colours and *C. pechuelii* haplotypes from Namibia are illustrated by striped patterns. The size of circles indicates frequency of haplotypes and actual haplotype frequencies are indicated in insert.

Figure 3: BEAST chronogram (based on *ITS* and *trnL-F*) for the genus *Cardiospermum*, including species with unknown native ranges *C. grandiflorum*, *C. corindum*, *C. halicacabum* and the African species *C. pechuelii*. Posterior probabilities higher than 0.9 are indicted at nodes and countries are indicated with different colours, following the key provided in figure 2. The scale bar indicates estimated time since divergence in millions of years (MYA).

Figure 4: MrBayes cladogram (based on *ITS*) for the genus *Cardiospermum*, including species with unknown native ranges *C. grandiflorum*, *C. corindum*, *C. halicacabum* and the African species *C. pechuelii*. Countries are indicated with different colours, following the key provided in figure 2. Posterior probabilities are indicated at nodes.

Chapter 3

Figure 1: Typical morphological and environmental characteristics of *Cardiospermum corindum* (A-C) and *C. pechuelii* (D-F) in their natal ranges in Namibia: leaf morphology (A and D); fruit morphology (B and E); and habitats occupied: the savannah Waterberg region (C) and rocky outcrops of the Namib Desert (F).

Figure 2: Sampled localities of populations of *C. corindum* in southern Africa (orange) and *C. pechuelii* (green) in Namibia and *C. corindum* in South America (red) (left). Population genetic structure was estimated from individuals from these populations using Bayesian assignment tests. For *K*=2 (number of genetic clusters) each horizontal bar illustrates a different individual and each colour the proportion of its genome assigned to each genetic cluster (right).

Figure 3: Bayesian phylogeny based on two cpDNA (*rpl32* and *trnL-F*) and one nDNA (*ITS*) gene sequences indicating the relationships among *Cardiospermum corindum* from South America and southern Africa and *C. pechuelii* from Namibia (left hand side). Branch support is given as posterior probabilities and accessions colour coded according to species and geographic affiliations. A haplotype network (right hand side) shows relationships for these species based only on the cpDNA. The sizes of the circles in the network are proportional to the amount of individuals sharing each haplotype and actual haplotype frequencies are indicated in insert.

Figure 4: Neighbor-net network based on genetic distances estimated from AFLP data for *Cardiospermum corindum* from South America (red) and *C. pechuelii* and *C. corindum* from southern Africa (yellow). Only bootstrap support >70% are shown in the network.

List of Tables

Chapter 1

Chapter 2

None

Chapter 3

heterozygosity (*He*) and % polymorphic Loci (*p*). C) Pairwise Population Matrix of Nei Genetic Distance.

Supplementary information

Chapter 1

Figure S1: Response plots of the variables used for modelling. Response plots for A) *C. corindum* global data, B) *C. corindum* native data, C) *C. grandiflorum* global data, D) *C. grandiflorum* native data, E) *C. halicacabum* global data and F) *C. halicacabum* native data. The plots are arranged according to the importance in brackets with the presence (top) and absence (bottom) points shown on the graph.

Figure S2: Species distribution modelling predictions. Modelling predictions for A) *C. corindum* B) *C. grandiflorum* and C) *C. halicacabum* using native range presence data and pseudo absence points generated with global range data.

Table S1. Contribution of each BioClim variable for distribution modelling of three *Cardiospermum* species.

Table S2. Evaluation of modelling predictions. True skill statistic (TSS) and area under the ROC curve (AUC) for native range presence data and global pseudo absence data, with independent data evaluation for the native range models.

Chapter 2

Table S1. Sample ID, species, country, latitude and longitude of samples used in this study.

Figure S1: BEAST chronogram (based on *ITS* and *trnL-F*) with an exponential distribution at the root calibration, thus enforcing a younger tree for the genus *Cardiospermum*, including species with unknown native ranges *C. grandiflorum*, *C. corindum*, *C. halicacabum* and the

African species *C. pechuelii*. 95% confidence intervals for nodal ages are shown and scale bar indicating estimated time since divergence in millions of years.

Chapter 3

Table S1. Sample ID, species, latitude and longitude of samples used in this study.

Figure S1: Structure harvester results using the Evanno Method indicating the optimum number of genetic clusters (K) as 2.

Thesis outline

Invasive species directly impact biodiversity and often lead to economic losses and it is therefore crucial to effectively manage them. Species of balloon vines within the genus *Cardiospermum* have been extensively moved around the world as ornamental and medicinal plants and have since escaped plantations becoming naturalised in many parts of the world. Two species have near cosmopolitan distributions, *C. grandiflorum* and *C. halicacabum*, while a third species, *C. corindum*, has a widespread distribution in the neotropics and Africa. At least one species (*C. grandiflorum*) is an accepted invasive species in southern Africa and other parts of the world, while the historical biogeography of *C. halicacabum* and *C. corindum* is problematic in many regions globally. This in turn hampers ongoing and effective management of invasive populations. For example, the unknown native statuses of *C. corindum* and *C. halicacabum* in southern Africa have so far prevented the release of already-identified biological control agents due to potential non-target impacts.

Against this background I reviewed the genus with focus on the most widespread species, *C. halicacabum* and *C. grandiflorum*, by conducting an exhaustive literature search (Chapter one). I report on their introductions histories, biology and ecology while also discussing different management options. I reconstructed a nuclear gene phylogeny for the most widespread species in the genus to shed light on their biogeographic uncertainties within southern Africa. This chapter also reports on potential suitable habitats globally for the three most widespread species, *C. grandiflorum*, *C. halicacabum* and *C. corindum*, using species distribution modelling. From these results I tested the accuracy with which this commonly employed method can predict known non-native ranges using native range data to predict current proposed alien ranges.

In the second chapter I address the biogeographic uncertainties within the genus *Cardiospermum*, resolving unknown native ranges, adding to the preliminary results of chapter one. More specifically, I use three genetic markers to reconstruct a chloroplast haplotype network and dated multi-gene phylogeny to infer potential native ranges for *C. halicacabum* and *C. corindum* in southern Africa and other parts of the world by assessing the evolutionary relationships within and among selected balloon vine species.

In chapter three I investigate the possibility of adaptive evolution of the Namib Desert endemic species, *C. pechuelii*, following long distance dispersal of *C. corindum* from South America to Africa. In chapter two I determined that *C. corindum* arrived on the African shore using long distance dispersal, evident from a monophyletic clade and close relationship between *C. corindum* from southern Africa and *C. pechuelii* compared to South American *C. corindum*. Long distance dispersal is characterised by reduced genetic variation known as the founder effect and along with novel environmental conditions, can lead to rapid diversification. Again, using three different genetic markers I reconstruct a chloroplast haplotype network and multi-gene phylogeny to infer the evolutionary history of African *C. pechuelii* and *C. corindum*. Moreover, to investigate population genetic structure and diversity between *C. corindum* and *C. pechuelii* populations, I use amplified length polymorphism markers (AFLPs) for sampled *C. corindum* and *C. pechuelii* populations.

I summarize all my research results and their implications for invasive species management in chapter four. I focus primarily on the release of already-identified biological control agents against invasive *C. grandiflorum* in South Africa and briefly discuss potential future studies.

1. The ecology, biogeography, history and future of two globally important weeds: *Cardiospermum halicacabum* L. and *C. grandiflorum* Sw.

Accepted for publication in *Neobiota* (08 August 2013)

1.1 Abstract

Members of the balloon vine genus, *Cardiospermum*, have been extensively moved around the globe as medicinal and horticultural species, two of which are now widespread invasive species; *C. grandiflorum* Sw. and *C. halicacabum* L. A third species, *C. corindum* L., may also have significant invasion potential. However, in some regions the native status of these species is not clear, hampering management. For example, in South Africa it is unknown whether *C. halicacabum* and *C. corindum* are native, and this is a major constraint to on-going biological control programmes against invasive *C. grandiflorum*. We review the geography, biology and ecology of selected members of the genus with an emphasis on the two most widespread invaders, *C. halicacabum* and *C. grandiflorum*. Specifically, we use molecular data to reconstruct a phylogeny of the group in order to shed light on the native ranges of *C. halicacabum* and *C. corindum* in southern Africa. Phylogenetic analyses indicate that southern African accessions of these species are closely related to South American taxa indicating human-mediated introduction and/or natural long distance dispersal. Then, on a global scale we use species distribution modelling to predict potential suitable climate regions where these species are currently absent. Native range data were used to test the accuracy with which bioclimatic modelling can identify the known invasive ranges of these species. Results show that *Cardiospermum* species have potential to spread further in already invaded or introduced regions in Australia, Africa and Asia, underlining the importance of resolving taxonomic uncertainties for future management efforts. Bioclimatic modelling predicts Australia to have highly favourable environmental conditions for *C. corindum* and therefore vigilance against this species should be high. Species distribution modelling showed that native range data over fit predicted suitable ranges, and that factors other than climate influence establishment potential. This review opens the door to better understand the global

biogeography of the genus *Cardiospermum*, with direct implications for management, while also highlighting gaps in current research.

Key words: balloon vines; biological invasion; *C. corindum*; management; phylogeny; species distribution modelling

1.2 Rationale

Understanding the biology, ecological requirements, and native distributions of potentially invasive species is crucial to ensure effective management and to predict their potential invasiveness. We review these attributes for selected members of a globally weedy genus, *Cardiospermum*, commonly known as balloon vines. We review the ecology and history of anthropogenic range expansion of the genus, with special emphasis on the two most problematic species in the group, *C. grandiflorum* and *C. halicacabum*. On a regional scale we aim to resolve the native provenance(s) of balloon vine species found in southern Africa, using a phylogenetic approach. Lastly, on a broad scale we assess the invasion risk posed by balloon vine species found outside their supposed native ranges, using species distribution modelling. Moreover, to evaluate the merit of this commonly employed method, we compare data of known invaded areas to predictions based on native range records.

1.3 Biogeography and phylogeny of selected *Cardiospermum* taxa

The genus *Cardiospermum* L. 1753 (family Sapindaceae, tribe Paullinieae) currently consists of 17 shrub, subshrub, climber, and erect species, commonly called balloon vines (Subramanyam *et al*. 2007). Around half of the species occur in moist tropical and subtropical regions while others are arid-adapted (Ferrucci & Urdampilleta 2011). Thirteen *Cardiospermum* species *(C. oliveirae, C. urvilleoides, C. procumbens, C. pterocarpum, C. anomalum, C. pygmaeum, C. cristobaliae, C. tortuosum, C. bahianum, C. integerrimum, C. heringeri, C. cuchujaquense, C. dissectum)* are mostly restricted in and around the Neotropics from south-eastern Brazil to north-central Mexico (Ferrucci & Umdiriri 2011) with most found in Brazil (12 spp.). Nine species are restricted to Brazil while the remaining eight species display wider geographical distributions. *Cardiospermum pterocarpum* occurs in Brazil,

Argentina and Paraguay. *Cardiospermum pygmaeum, C. dissectum* and C*. cuchujaquense* are restricted to Mexico with *C. dissectum* also having been recorded in Texas, USA. *Cardiospermum pechuelii* is the only taxon restricted to Africa, occurring only in the desert areas of Namibia. Three species, *C. corindum*, *C. halicacabum* and *C. grandiflorum* have near cosmopolitan distributions (Ferrucci & Umdiriri 2011, Urdampilleta *et al*. 2012).

Morphology divides this genus into three sections; *Cardiospermum* Radlk., *Carphospermum* Radlk. and *Ceratadenia* Radlk. (Urdampilleta *et al*. 2012). In addition to *Cardiospermum*, Paullinieae includes five other genera, *Serjania*, *Paullinia*, *Urvillea*, *Houssayanthus* and *Lophostigma*, of which *Urvillea* is regarded the sister genus to *Cardiospermum* (Ferrucci & Acevedo-Rodrigues 1998).

Only four *Cardiospermum* species occur abundantly outside the neotropics: *C. halicacabum*, *C. grandiflorum*, *C. corindum*, and *C. pechuelii* (Burke 2003, Ferrucci & Umdiriri 2011). *Cardiospermum pechuelii* may be the only true African taxon, found in the Namib Desert (Burke 2003, Simelane *et al*. 2011). *Cardiospermum pechuelii* is morphologically similar to other arid adapted species, such as *C. dissectum* from Mexico. The most widely distributed species are tropical and subtropical *Cardiospermum corindum* (Fig. 1A), *C. grandiflorum* (Fig. 1B) and *C. halicacabum* (Fig. 1C) (Mc Kay *et al.* 2010, Simelane *et al.* 2011). All three species occur in the Neotropics and subtropical southern Africa. *Cardiospermum corindum* is also found in parts of India where it is known under its synonym name *C. canescens* (The Plant List 2010, Raju *et al*. 2011). *Cardiospermum grandiflorum* and *C. halicacabum* are present in Australia and other Pacific islands classified as alien or invasive, and *C. halicacabum* is also present in Europe and Asia (Subramanyam *et al*. 2007). In many of these countries the native status of these species is highly debated and their biogeographical history remains uncertain (Table 1). *Cardiospermum grandiflorum*, *C.*

corindum and *C. halicacabum* are regarded as being native in South and Central America while the status of *C. halicacabum* is questioned in North America (Henry & Scott 1981, Bowen *et al*. 2002, Carroll 2007, Goosem 2008) and tropical Africa (USDA, United States Department of Agriculture; Weeds of Australia). Similarly the status of *C. corindum* is uncertain throughout the continent African continent (Henderson 2001, Simelane *et al*. 2011). In Asia *C. halicacabum* is variously regarded as either alien or native (Venkatesh & Krishnakumari 2006, Subramanyam *et al*. 2007).

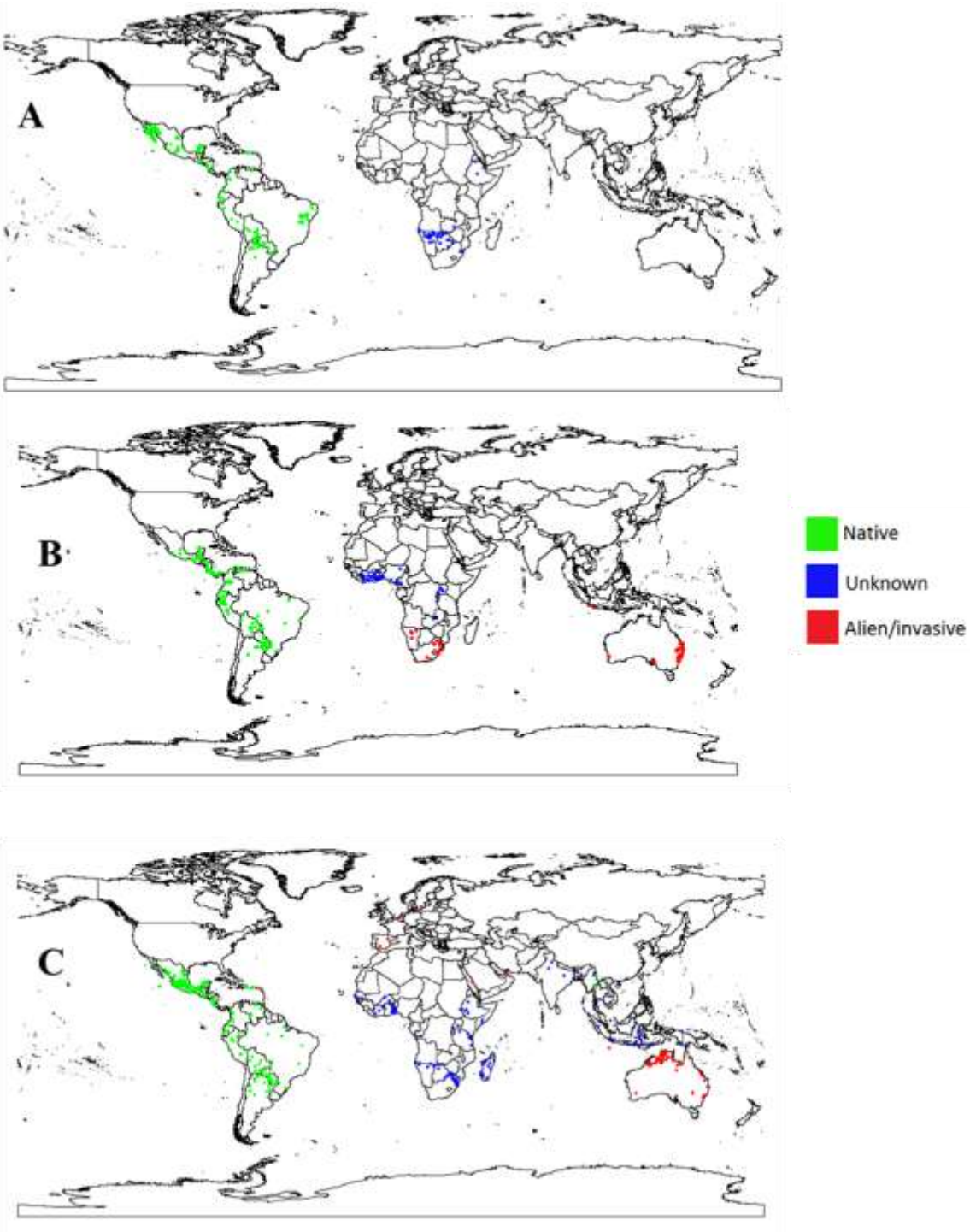


Figure 1: Distribution of *Cardiospermum* species. Global distribution of A) *C. corindum*, B) *C. grandiflorum* and C) *C. halicacabum* in native, unknown and alien or invasive regions.

Alien invasive species are a global concern and a threat to biodiversity (Pimentel *et al*. 2000, Van Wilgen *et al.* 2001). They also negatively impact agricultural and forestry sectors with substantial economic costs associated with their direct impacts, eradication, control and restoration efforts (Pimentel *et al*. 2000, 2001). Like many invasive species, *Cardiospermum* species have been introduced for their economic value prior to becoming problematic (Pimentel *et al.* 2000, Van Wilgen *et al.* 2001). *Cardiospermum* species have been extensively moved around the world for both their medicinal (Venkatesh Babu & Krishnakumari 2006, Subramanyam *et al.* 2007) and ornamental (Carroll *et al.* 2005a) values.

The ornamental attraction of *Cardiospermum* species are their inflated balloon shaped fruit (Fig. 2). Coincidently this trait also contributes to their colonisation success, since these balloons can float in seawater and stay viable for long periods of time, facilitating long distance dispersal, even between landmasses (Carroll *et al.* 2005a, Simelane *et al*. 2011). For example, *C. grandiflorum* was introduced to the Cook Islands as a result of a hurricane (Meyer 2004), whilst increased spread of balloon vines in Australia was associated with a major cyclone and subsequent flooding (Carroll *et al*. 2005a). We floated *C. grandiflorum* fruit structures in seawater and found some of them capable of floating more than 25 weeks with seed remaining viable. (E. Gildenhuys *et al*., unpubl. data). Upon dehiscence, each seed is attached to a circular blade that permits further transport by wind.

Figure 2: *Cardiospermum* fruit. The ornamental attraction of *Cardiospermum* plants and the reason for their widespread distribution is their balloon shaped fruit A) *C. grandiflorum* (E Gildenhuys) and B) *C. halicacabum* (J-Y Meyer).

Invasive *Cardiospermum* species are considered "transformer weeds" (Mc Kay *et al*. 2010), as they often extensively cover native vegetation, depriving it of sunlight and thus photosynthesis (Mc Kay *et al*. 2010, Simelane *et al.* 2011). *Cardiospermum* invasions also have substantial economic impacts on sugarcane and soybean production (Johnston *et al.* 1979, Jolley *et al*. 1983, Voll *et al.* 2004, Subramanyam *et al*. 2007, Murty & Venkaiah 2011). For example, in Brazil *C. halicacabum* reduces soybean crop yields by up to 26% (Dempsey *et al.* 2011, Brighenti *et al.* 2003). The problem with controlling *Cardiospermum* infestations in soybean crops is the difficulty of mechanically excluding their seeds, which are similar in size and shape to those of soy (Brighenti *et al.* 2003).

1.3.1 Two balloon vine species well-travelled

Currently two *Cardiospermum* species are globally considered important invaders. *Cardiospermum grandiflorum* is classified as an invasive species in Australia, southern Africa, Cook Islands and many other Pacific islands (Mc Kay *et al*. 2010) while *C. halicacabum* is considered a weed in Australia with its status (native or introduced) undetermined in most

other parts of its range (Henderson 2001, Harris *et al*. 2007). In Australia, *C. grandiflorum* is considered amongst the “most destructive life forms of rainforests” (Werren 2002), while in South Africa *C. grandiflorum* is classified as a Category 1 weed which means it’s cultivation is prohibited and control is mandatory (Henderson 2001).

South Africa’s Working for Water program launched a research initiative in 2003 to find biological control agents against *C. grandiflorum* (Simelane *et al*. 2011). Eight insects and two fungal agents have been identified and are currently undergoing host-specificity testing in South Africa (Simelane *et al.* 2011). Most are capable of feeding and developing on other *Cardiospermum* spp. in South Africa, in particular *C. halicacabum* and *C. corindum* (Mc Kay *et al*. 2010). Three promising agents were identified, a seed-feeding weevil (Curculionidae: *Cissoanthonomus tuberculipennis*), a fruit-galling midge (Cecidomyiidae: *Contarinia* spp.) and the rust fungus *Puccinia arechavaletae* (Simelane *et al*. 2011). Concerns about potential non-target impacts of candidate control agents on *C. corindum* and *C. halicacabum,* as well as the debated native status of these congeners in southern Africa (Table 1), have so far prevented the release of these agents.

Table 1. Details of uncertain native or non-native statuses of two *Cardiospermum* species in North America and Africa.

	Continent	References for debated native/non-native status
C. halicacabum	North America	Brizicky 1963; James, 1825; Carroll & Boyd 1992
	Africa	Brizicky 1963; Davies & Verdcourt 1998; Hyde & Wursten 2012a, b; Henderson 2001; Llewellyn

		et al. 2008; Simelane *et al*. 2011
C. corindum	Africa	Davies & Verdcourt 1998; Henderson 2001; Simelane *et al*. 2011; Germishuizen *et al*. 2006; Adeyemi & Ogundipe 2012
	North America	Brizicky 1963, Castellanos *et a*l. 1999, Molina-Freaner and Tinoco-Ojanguren 1997

1.3.1.1 Invasion histories of *C. grandiflorum* and *C. halicacabum*

The ornamental trade of *Cardiospermum halicacabum* and *C. grandiflorum* spans more than 100 years. For example, in Australia the first herbarium records of *C. grandiflorum* date back to 1923, collected around Sydney, New South Wales (Carroll *et al.* 2005a). Currently invasive populations are found throughout the east coast of Australia between Sydney and Cairns although less abundantly to the north of Brisbane (E. Gildenhuys, pers. obs.). More recently the species has spread inland to forest areas such as Toowoomba (Queensland) and the Blue Mountains (New South Wales) (Carroll *et al*. 2005a, E. Gildenhuys, pers. obs.). *Cardiospermum halicacabum* is more abundant in the northern parts of Australia such as Darwin and Cairns, and is seldom found along the east coast south of Rockhampton, Queensland (E. Gildenhuys, pers. obs.). It is speculated that *C. halicacabum* was introduced during James Cook's second voyage in the 1770's long before the introduction of *C. grandiflorum* (Bean 2007, Harris *et al*. 2007).

The introduction of *Cardiospermum grandiflorum* into South Africa occurred approximately 100 years ago (Simelane *et al*. 2011). Today it is classified as a major weed, and is present and considered invasive in five provinces, of which Kwazulu-Natal and the Eastern Cape are the most affected (Henderson 2001, Simelane *et al*. 2011). The first

records of *C. halicacabum* in South Africa dates back to 1917, 1919 in Namibia and 1930 in Botswana (Global Biodiversity Information Facility: GBIF, http://data.gbif.org/welcome.htm). It is classified as a minor weed in southern Africa, though its native status is debated, with slight impacts compared to *C. grandiflorum* (Henderson 2001).

Cardiospermum halicacabum and *C. grandiflorum* are also present in North America (Carroll *&* Loye. 2012). *Cardiospermum halicacabum* is more widespread than *C. grandiflorum*, the latter apparently restricted to a small area in suburban Los Angeles (S. Carroll, pers. obs.). Due to the evident ability of some *Cardiospermum* species to disperse over long distances (Carroll *et al.* 2005a, Simelane *et al*. 2011), it is possible that the presence of *C. halicacabum* in North America is due to natural dispersal from South and Central America, rendering a native status. On the other hand, if seeds escaped horticultural and agricultural environments, they should be awarded non-native status (Subramanyam *et al*. 2007). *Cardiospermum halicacabum* was reported in the Spontaneous Illinois Vascular Flora before 1922 and was described as abundant in Oklahoma in the 1820's (James 1825); thus, if not native, *C. halicacabum* was introduced more than 180 years ago.

Cardiospermum halicacabum is also present in China and India. In China it is described as a common weed in forest margins, shrublands, grasslands, cultivated areas and wastelands of the east, south and southwest (Flora of China, www.eFloras.org) – though considered native by some – [Pacific Island Ecosystems at Risk (PIER)]. In India it is widespread and considered non-native (Raju *et al*. 2011). The history of *C. halicacabum* in these countries is unknown, but it is widely used for medicinal purposes (Subramanyam *et al*. 2007).

1.3.1.2 Biology and ecology of *C. grandiflorum* and *C. halicacabum*

A comprehensive understanding of the biology and ecology of *C. halicacabum* and *C. grandiflorum* is important because of the invasive potential and biogeographic uncertainties which characterise these two taxa. Such information will also contribute to making informed decisions on their conservation (if native) or control (if invasive). This is especially true since the extent to which these species are invasive is essentially unknown and the uncertainties of their classification in most areas suggest the possibility of a cosmopolitan native distribution.

The morphology of these two species is similar, with both being adapted for tropical and subtropical climates. *Cardiospermum grandiflorum* is a large, semi-woody perennial, whereas *C. halicacabum* is smaller, less woody and commonly annual. *Cardiospermum grandiflorum* has elongated fruit (4.5-6.5 cm in length) compared to the more compact fruit of *C. halicacabum* (2.5-3.0 cm in length) (Fig 2A & B). Fruit structures consist of three dorsally keeled membranous capsules each consisting of three internal blades (Weckerle & Rutishauser 2005). The fruit are septifragal with the capsules breaking away from each other when fruit are ripe, changing colour from green to brown (Weckerle & Rutishauser 2005). Seeds of the two species differ, with a kidney shaped hilum on *C. halicacabum* seeds and a round hilum on *C. grandiflorum* seeds. Both species normally produce three seeds per fruit (Weckerle & Rutishauser 2005), are climbers with tendrils and have large flat biternate leaves. The leaves and stems of *C. grandiflorum* have small reddish hairs that are absent in *C. halicacabum* (Henderson 2001). Flowers are white and yellow with *C. halicacabum* flowers smaller (2-3 mm) compared to those of *C. grandiflorum* (7-11 mm) (Henderson 2001). The average length of *C. halicacabum* is 1-3 m, while *C. grandiflorum* is slightly taller with an average of 2-5 m, though both are capable of greatly exceeding these lengths (Henderson 2001).

Both taxa produce flavone aglycones and cyanogenic compounds that likely protect them against predators such as soapberry bugs (Subramanyam *et al*. 2007). Soapberry bugs (genera *Leptocoris, Jadera* and *Boisea* from the family Rhopalidae) feed exclusively on seeds of Sapindaceae and are predators of *Cardiospermum* (Carroll *et al.* 2005b, Carroll 2007). An example of the impact of invasive *Cardiospermum* populations includes an evolved increase in beak length of the native *Leptocoris tagalicus* soapberry bug feeding on invasive *C. grandiflorum* in Australia (Carroll *et al.* 2005b). Soapberry bugs co-occur with the widespread distribution of *Cardiospermum* and thus may be a factor in *Cardiospermum* reproduction globally. A treatment of soapberry bugs that feed on *C. halicacabum* and *C. grandiflorum* can be found in Carroll & Loye (2012).

The germination and growth success of *Cardiospermum halicacabum* is well studied because of its medicinal value, as well as its impact on soybean plantations and on natural riparian areas (Dempsey 2011). In contrast, no studies exist addressing these topics for *C. grandiflorum,* despite the need for additional biological information about this environmental weed. Optimum germination of *C. halicacabum* takes place at 35°C, with high oxygen concentrations increasing germination success (Johnston *et al*. 1979, Jolley *et al*. 1983, Dempsey 2011). Therefore, in natural habitats, establishment may be more likely in conditions with warm, well-oxygenated soils. Seeds and young plants are able to survive flooded, saturated and dry conditions while performing best in intermediate conditions (Dempsey 2011).

Despite morphological similarity, these two species differ markedly. They occasionally occur sympatrically but mostly prefer different habitats with *C. halicacabum* dominating tropical and *C. grandiflorum* subtropical areas (Henderson 2001). Although both species invade forest margins and watercourses, *C. grandiflorum* also thrives in disturbed urban open

areas while *C. halicacabum* predominantly invades wood- and grasslands which highlights its threat to plantations (Henderson 2001).

1.3.1.3 Management of invasive *Cardiospermum*

To date, managing and reducing impacts of *Cardiospermum* invasions has mostly involved manual removal or burning (Subramanyam *et al*. 2007). Manual removal involves cutting plants at the base enabling the top part to die off after which roots are dug out which is thus labour intensive (Mc Kay *et al*. 2010). Chemical control of larger plants includes treatment with paraquat, glufosinate-ammonium, lactofen, carfentrazone-ethyl, sulfentrazone, glyphosate or 2, 4-Dichloraphenoxy acetic acid (Subramanyam *et al*. 2007). However, the use of chemical control could potentially be problematic for two reasons, firstly because of non-target impact on underlying vegetation and secondly the typical proximity of invasions to waterways makes environmental contamination a threat (Simelane *et al*. 2011). Another key problem in the management of *Cardiospermum* invasions is the persistent seed bank, with seeds able to survive for up to eleven years (FloraBase 2012, Save Our Waterways Now 2013). If the weedy canopy is cleared it opens the door for long-lived seeds to sprout (FloraBase 2012).

1.3.1.4 Management and problems in South Africa

In collaboration with South Africa's Working for Water program, a biological control programme was initiated against *C. grandiflorum* in 2003. However due to the taxonomic uncertainty surrounding *C. halicacabum* and *C. corindum* (discussed earlier, Table 1), biocontrol agents cannot be released, hampering effective management in South Africa. The importance of clarifying the geographic native ranges of all *Cardiospermum* species currently found in South Africa for the successful biological control of *C. grandiflorum* is therefore

evident. If *C. corindum* and *C. halicacabum* are indeed native to southern Africa, only agents that are specific on *C. grandiflorum* can qualify for release in South Africa, and thus far, these agents have proved particularly difficult to rear and test under quarantine conditions (D. Simelane, pers. comm.). On the other hand, if *C. corindum* and *C. halicacabum* are not native to southern Africa, all suitable agents against *C. grandiflorum* qualify for release in South Africa.

1.4 Molecular systematics of C*ardiospermum* species in southern Africa

To determine the relationship between *Cardiospermum* species occurring in Africa and South America we sequenced two accessions of *C. grandiflorum*, *C. halicacabum* and *C. corindum* from each continent (South America and Africa). DNA was extracted from dried plant material using the CTAB method (Doyle & Doyle 1990). The internal transcribed spacer gene region was amplified using primers ITS1 and ITS4 (For more information regarding specific primers and PCR conditions see chapter 2). A phylogenetic tree was then reconstructed in BEAST version 17.4 (Drummond *et al*. 2012) using a General Time-Reversible (GTR + G) model with uneven rates of evolution between base pairs.

The retrieved phylogeny indicates a close relationship between samples from South America and southern Africa (Fig. 3). For *C. grandiflorum* and *C. halicacabum* southern African samples are more closely related to South American samples than to other samples from southern Africa (i.e. geographic paraphyly). It is therefore likely that *C. halicacabum* in southern Africa, like *C. grandiflorum*, represents a recent introduction, and is therefore not native. For *C. corindum* however the phylogeny cannot dismiss natural long distance dispersal as an explanation for the species' presence in southern Africa, due to the southern African accessions forming a monophyletic group within the South American clade. The ability

of *Cardiospermum* fruit to float in seawater for long periods of time and remain viable, makes a strong case for long distance dispersal. In order to clarify the uncertainty around human introduction versus rare long distance dispersal events, future phylogenetic analyses should include more and geographically widespread collections.

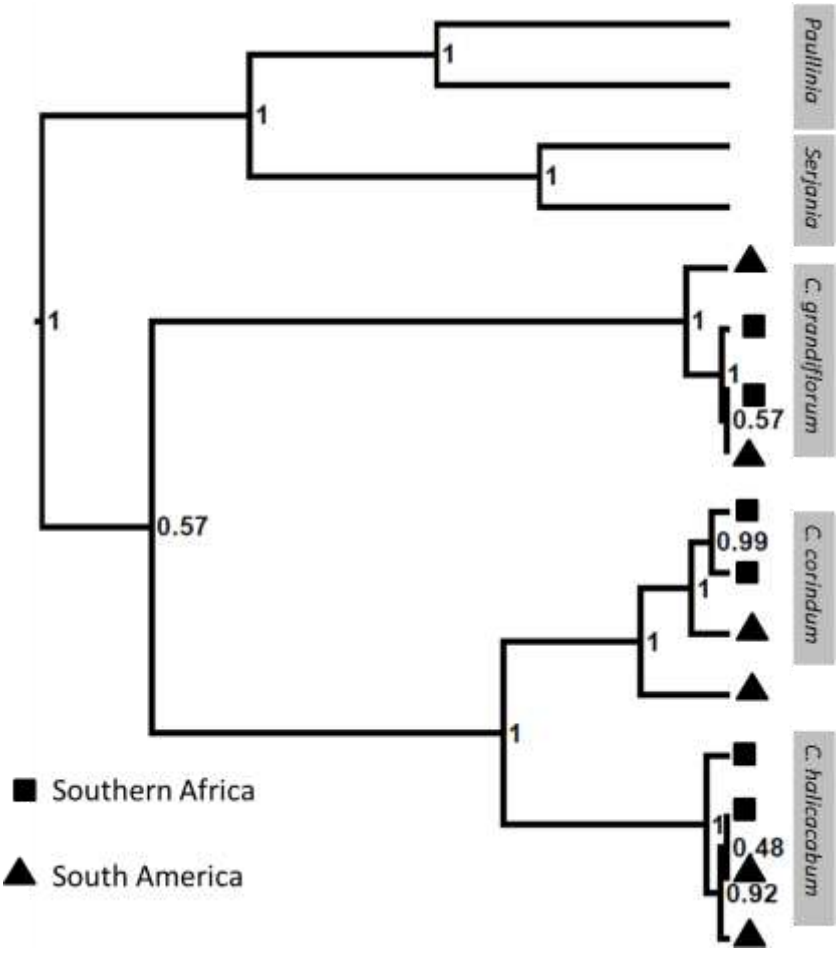


Figure 3: *Cardiospermum* phylogeny. Phylogeny of six South American and six southern African accessions of *Cardiospermum* species with *Paullinia* and *Serjania* species used as outgroup taxa. Topology support is shown as posterior probability at each node.

1.5 Bioclimatic preferences of *Cardiospermum halicacabum*, *C. grandiflorum* and *C. corindum*

Prevention is better than cure, with eradication of introduced species typically becoming less feasible as spread progresses (Thuiller *et al.* 2005). Identifying a species' suitable climatic range can therefore help to determine areas where introduction should be prevented or management intensified. Species distribution modelling is probably the most popular method for determining such areas (Allouche *et al.* 2006, Hirzel *et al.* 2006). Essential to the accuracy of species distribution modelling is the assumption that niche shifts do not occur in a newly introduced area, which has been shown to occur rarely (Petitpierre *et al*. 2012).

1.5.1 Modelling methods

We used BIOMOD version 1.1.5 (Thuiller *et al.* 2009) implemented in R version 2.15.1 (R Development Core Team 2012) to predict potentially suitable climate habitats for *C. halicacabum*, *C. grandiflorum* and *C. corindum*. Locality records were sourced from public databases [GBIF; Henderson 2007] and personal observations. We discarded records with spatial uncertainty (e.g., points in the ocean) and those from botanical gardens or with missing or duplicate values. Since no absence data is available for *Cardiospermum* species, but is needed for modelling, 10,000 pseudo-absence background points were created per species, by random sampling of the Köppen-Geiger climate classification. We employed generalized boosted regression models (GBM), a method uniting regression trees with boosting (for a more comprehensive description see Elith *et al.* 2008). For all analyses, seven climatic variables were sourced from BioClim (Hijmans *et al*. 2005), based on their importance for species survival and low co-linearity (Table 2). Importance, and thus the contribution of each variable to the model was assessed using Pearson rank correlation between standard

predictions and those based on random permutations for each variable separately (Thuiller *et al*. 2009). If correlations between these two predictions were high, the specific variable was regarded as less important. Co-linearity between different variables was limited to <0.70 using Spearman rank correlation coefficients. Consequently, precipitation of the wettest quarter was dropped for modelling of *C. halicacabum* due to a high correlation with precipitation of the warmest quarter. A raster of 6 arc min was used to extract variables since a more coarse resolution is realistic for global scale prediction, while also accounting for sampling error. Models were calibrated with 70% of the data and evaluated with the remaining 30%. A cut-off value was determined with BIOMOD's default setting, representing the best probability threshold which maximizes the percentage of presence and absences correctly predicted for the evaluation data (Thuiller *et al*. 2009). Area under the receiver-operator-curve (AUC, Hanley & McNeil 1982) and the true skill statistic (TSS, Allouche *et al.* 2006) were used for model evaluation. AUC scores between 0.95 and 1 indicate an excellent, 0.9 and 0.95 a good and 0.6 and 0.8 a fair model (Thuiller *et al.* 2005). TSS values of 0.8-1 are excellent, 0.6-0.8 good and 0.0-0.6 fair for predicting accuracy (Allouche *et al.* 2006).

Table 2. Contribution (%) of each BioClim variable used for distribution modelling of *Cardiospermum* species. The first value in each species column is for global and the second for native range modelling.

	Variable importance					
Variables used for modelling	*C. halicacabum*		*C. grandiflorum*		*C. corindum*	
	Global	Native	Global	Native	Global	Native
Min temperature of the coldest month	21.2	12.5	13.8	25.4	14.9	21.1
Max temperature of the warmest	6.2	2.3	4	0.9	3.9	1.7

month						
Precipitation of the coldest quarter	4.9	22.2	27.8	25.9	13.7	2.1
Precipitation of the driest month	2	1.1	13.1	2.5	3.6	16.9
Precipitation of the warmest quarter	44.5	8.3	20	22.1	31.5	2.7
Temperature seasonality	17.2	57.6	22.8	24.9	6	7.9
Precipitation of the wettest quarter	-	-	3.2	5.7	34	42.5

The accuracy of species distribution modelling is influenced by false positives and negatives (Thuiller *et al.* 2005, Fawcett 2006). Therefore a second aim of our species distribution modelling approach was to evaluate the accuracy with which this technique can predict potential invasive regions using models calibrated with native range data only. South and Central America were used as the native range for all three species since native status is debated in all other regions. A model calibrated using these records were then used to project suitable climate regions globally as described above. Known global occurrence records were then used as independent data to evaluate modelling accuracy.

1.5.2 Modelling results

Australia: Global data models for all three species performed well, with AUC values above 0.9 and TSS values above 0.65 (Table 3). Bioclimatic predictions show that a large proportion of Australia is climatically suitable for *Cardiospermum corindum*, a species currently absent in this country. Both *C. halicacabum* and *C. grandiflorum* have been introduced to Australia and are classified as invasive weeds. The suitable climate range for *C. corindum* in Australia is much larger than predicted for both *C. grandiflorum* and *C. halicacabum* and as such ornamental or medicinal introductions of *C. corindum* into Australia

should be prevented (Fig. 4A, B & C). Modelling also predicted that the east coast of Australia is climatically highly suitable for *C. halicacabum,* such that any risks from its establishment in this area should be assessed. *Cardiospermum grandiflorum* appears to be a more rapid colonizer than *C. halicacabum* in Australia and it is already present in most predicted areas. It is however likely to become locally more abundant in areas where it is already found (Fig. 1B & Fig.4B).

Table 3. Evaluation of modelling predictions. True skill statistic (TSS) and area under the ROC curve (AUC) for global and native range modelling of three widespread *Cardiospermum* species. The first value in TSS and AUC column is for global and the second for native range modelling. Independent data evaluation is for the native range models evaluated against known non-native ranges.

Species	TSS		Independent data (TSS)	AUC		Independent data (AUC)
	Global	Native	Native	Global	Native	Native
C. halicacabum	0.651	0.703	0.441	0.9	0.923	0.755
C. grandiflorum	0.759	0.665	0.343	0.95	0.895	0.639
C. corindum	0.689	0.629	0.565	0.905	0.896	0.881

Europe and Asia: Our modelling approach identified Europe as mostly climatically unsuitable for *Cardiospermum* (Fig. 4A, B & C). Areas of suitable climate are present for all three species in certain parts of Asia including India (where *C. halicacabum* and *C. corindum* are present), Thailand and Pakistan, with *C. grandiflorum* potentially being the most restricted

taxon (Fig. 4B). *Cardiospermum corindum* has high climatic suitability in southern Yemen, southern India, Thailand, Myanmar and southern China (Fig. 4A). The southernmost tip of Yemen seems climatically suitable for *C. halicacabum*, with India, Thailand, Cambodia, Vietnam, Myanmar, Japan, Taiwan and parts of China highly suitable (Fig. 4C). Many of these regions are already occupied by *C. halicacabum*. Climatically suitable habitat for *Cardiospermum grandiflorum* in Asia only appears to be present in southern India, Sri Lanka and parts of Vietnam (Fig. 4B).

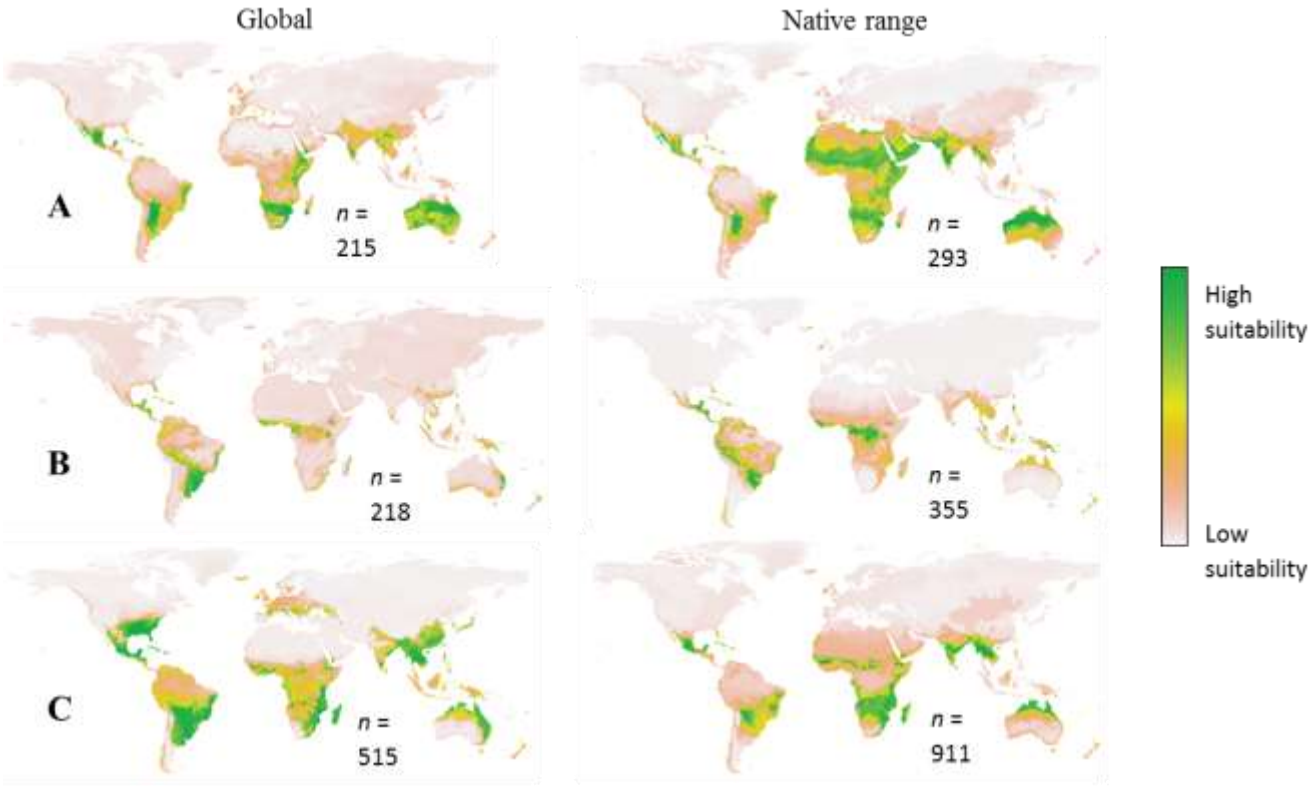


Figure 4: Species distribution modelling of *Cardiospermum* species. Global climatically suitable ranges for A) *C. corindum* B) *C. grandiflorum*, and C) *C. halicacabum* as predicted by

boosted regression trees in BIOMOD using global (left) and native range data (right). Number of occurrence points used for modelling (*n*) is indicated on each map.

Southern Africa*:* In South Africa bioclimatically suitable areas for *C. grandiflorum* are in the Western Cape Province, while for *C. halicacabum* they are in coastal areas in the Eastern Cape Province. Bioclimatically suitable areas in South Africa are the largest for *C. corindum,* with the Western and Eastern Cape Provinces being highly suitable. Currently the species is limited to Limpopo, Mpumalanga and northern parts of Kwazulu Natal (SANBI). Spread and anthropogenic movements of *Cardiospermum* species in South Africa should therefore be closely monitored since a large part of South Africa appears climatically suitable for establishment. While *Cardiospermum grandiflorum* and *C. halicacabum* are recorded as naturalised in parts of Namibia and Botswana, bioclimatic modelling did not predict either country as climatically suitable. *Cardiospermum* species are not widespread in these two countries and possibly only occur in areas with suitable microclimates. Such habitats typically differ significantly from surrounding environments and often result from human actions, and are therefore excluded in bioclimatic modelling based on more coarse data, such as this study (Kearney & Porter 2009).

1.5.3 Testing model accuracy

Models calibrated with South and Central American native occurrence records performed fairly well when cross-validated using AUC and TSS, with values higher than 0.85 and 0.6 respectively. However this was not the case when these models were evaluated with independent data, thus known presence data not used in modelling. *Cardiospermum halicacabum* and *C. grandiflorum* had low AUC and TSS values ranging between 0.60-0.80

and 0.30-0.45 respectively, only *C. corindum* models performed fair (AUC > 0.85 and TSS > 0.55, Table 3).

These results indicate that models calibrated with native range occurrence records only, would not have accurately predicted the invasive spread of *C. grandiflorum* in South Africa while underestimating its potential range in Australia. This lack of accuracy for identifying invasive regions using native data questions the suitability of using species distribution modelling alone when determining potential invasive regions.

Also contrary to what we expected, models calibrated using native range data predicted larger climatically suitable areas than models calibrated with global range data (Fig. 4; except for *C. halicacabum*). We hypothesised that this is due to the more restricted climate zones created with the widespread pseudo-absence data of the global range, thus including more diverse habitats to exclude as suitable areas. We plotted the presence and absence points for both native and global range data for each variable against the probability of occurrence using the response plot function in R version 2.15.1 (Fig. S1 A-F). In these figures it is clear that global data variables include a wider environmental range for pseudo-absences compared to the native range pseudo-absences, especially when considering the most significant variables based on variable importance (Table 2). To test if this is indeed the case we ran three additional models with the same settings as the previous models but using native range presence data and global pseudo-absences data. We used the same evaluation parameters as for the previous models (Table S1& S2). This approach resulted in projections that more closely resembled global range model predictions or are even more restricted predictions (Fig. S2). These results indicate that while native range data can be used to predict potential suitable areas, data are often over-fitted, thus over predict the extent of suitable habitats, due to less restricted absence data created from the native range.

1.5.4 Usefulness of bioclimatic species modelling

While species distribution modelling is a popular tool for predicting potential invasive ranges its accuracy remains questionable (Araújo & Luoto 2007, Sinclair *et al*. 2010). Bioclimatic modelling did not accurately predict current invasive regions for the widely naturalized species *C. grandiflorum*. Also native range data alone led to an over estimation of potential suitable habitats for *C. corindum* and *C. grandiflorum*. Our results comparing predictions based on native and global occurrence records are surprising and significant. We hypothesized that the reason for this observation is the more restricted climate zones created when using global pseudo-absences for model calibrations, an effect that can potentially be amplified for species characterised by incomplete range filling in their native ranges. A key assumption of species distribution modelling is pseudo-equilibrium, however this is probably unrealistic for most species and may therefore seriously impact model accuracy (Guisan & Thuiller 2005). On the other hand, bioclimatic predictions may be hampered if a species has undergone a niche shift in its invasive range (Broennimann *et al*. 2007). All the above-mentioned issues highlight how factors other than climate may play a crucial role in the accuracy of species distributions modelling. For example niche shift in the non-native range could be the result of release from natural enemies (Keane & Crawley 2002). Similarly, increased resource availability in the introduced range (Davis *et al*. 2000, Thompson *et al*. 2004) may increase habitat suitability while abiotic attributes of the new range may permit spread into novel habitats. In concert, dispersal limitations (Pulliam 2000), anthropogenic effects and unique historical factors (Jiménez-Valverde *et al*. 2008) may limit the distribution of species in their native ranges.

Thus, taking the contradicting results into account and also considering the many other factors that influences a species distributional range, lead us to conclude that while

bioclimatic modelling is a useful approach, it should not be used as a stand-alone tool when making conservation decisions regarding the introduction of species into a novel range and caution should be exercised to ensure the quality of input data while also taking other factors into account as discussed above.

1.6 Conclusions

Many regions globally appear climatically suitable for establishment of *Cardiospermum grandiflorum*, *C. corindum* and *C. halicacabum,* cautioning against further introductions. Resolving the native ranges for these species globally is therefore important for biodiversity conservation and invasive species management. For example, our preliminary results indicate that *C. halicacabum* from southern Africa have a close relationship with South American samples, but that rare long distance dispersal cannot be ruled out as an explanation, while the split between South American and southern African *C. corindum* hints towards a native status on both continents. Future work should include a more comprehensive phylogeny to substantiate our findings, including balloon vine specimens from other biogeographic regions where the native status is known. If it is found that they are indeed alien to Africa and Asia, a risk assessment challenge lies ahead since large areas of these continents appear climatically suitable for their establishment. No *Cardiospermum* species are regarded as native in Australia, and measures to limit the spread of *C. halicacabum* and *C. grandiflorum* may be augmented with biological control measures that include native soapberry bugs that are evolving to use them more efficiently (Carroll et al. 2005b). In addition, the introduction of *C. corindum* should be prohibited based on the wide environmental suitability identified for this species in Australia.

Cardiospermum species are also used by many people in rural areas for medicinal purposes, further emphasizing a need to resolve the natal biogeographic distribution of this globally important genus to ensure its effective management, control or conservation.

1.7 Acknowledgments

We thank Dr Ingolf Kühn and the two anonymous reviewers for their constructive comments on previous drafts of the manuscript. Financial support was provided by the DST-NRF Centre of Excellence for Invasion Biology and the Working for Water Programme through their collaborative project on “Research for Integrated Management of Invasive Alien Species”. E Gildenhuys acknowledges the South African National Research Foundation’s (NRF) Scarce Skills scholarship programme. J Le Roux also acknowledges Stellenbosch University’s Sub-committee B “Young Researchers Fund” and the NRF Thuthuka Programme for research funding. S Carroll acknowledges support from the School of Life Sciences at the University of Queensland, St. Lucia. We are grateful to Jason Donaldson and Vernon Visser for their help and advice with species distribution modelling.

2. Home sweet home: resolving biogeographic ranges of invasive balloon vines in the genus *Cardiospermum*

Submitted for consideration for publication to *Journal of Biogeography* (16 September 2013)

2.1 Abstract

Aim: A key aspect of effective invasive plant management is a clear understanding of natal ranges. In the balloon vine genus, *Cardiospermum*, uncertain native range biogeographies are hampering on-going biological control efforts against invasive taxa in southern Africa. With the objective of minimizing non-target impacts we aim to resolve unknown native ranges for widespread *Cardiospermum* species.

Location: Argentina, Australia, Brazil, Fiji, Hawaii, Mayotte, Namibia, South Africa, Tanzania, Tahiti and Uganda.

Methods: Phylogenetic relationships and dispersal histories of selected *Cardiospermum* species (*C. halicacabum*, *C. grandiflorum*, *C. corindum*) were inferred from a dated multi-gene phylogeny using a relaxed molecular clock and root age previously identified, constructed from the nuclear internal transcribe spacer region (*ITS*) and the non-coding chloroplast *trnL-F* intron. A second resolved Bayesian phylogeny was reconstructed for the *ITS* gene region representing more widespread samples. The chloroplast gene, *rpl32*, was also used to reconstruct a haplotype network to more closely investigate relationships within and among balloon vine species.

Results: Our phylogeny indicated polyphyly for *Cardiospermum halicacabum* worldwide, i.e. incongruence between geography and relatedness, with the chloroplast gene and haplotype network supporting these relationships. On the other hand, a deep divergence was identified for *C. corindum* from southern Africa and South America both within the haplotype network and phylogenetic tree (3.9 MYA). More surprising was the observed close phylogenetic relationship between African *C. pechuelii* and *C. corindum* supporting a native origin for *C. corindum* in southern Africa.

Main conclusions: We conclude that *C. halicacabum* is non-native and that both *C. corindum* and *C. pechuelii* are native to southern Africa. We confirm the invasive status of

C. grandiflorum in southern Africa but suggest the invasive status of this species be reviewed in other regions. We recommend that already-identified biological control agents not be released against invasive *Cardiospermum* taxa in South Africa due to potential non-target impacts that have previously been identified against native *C. corindum* and *C. pechuelii*.

Keywords: *Cardiospermum*; balloon vines; biological control; invasive; long distance dispersal; native range; plants; Phylogeography

2.2 Introduction

Understanding the historical processes that underlie current species distributions informs how regional biodiversity evolved, is maintained, and how it should be conserved (Cavender-Bares *et al.* 2009). Inferring these processes can be problematic for widespread species, especially for those having non-native distributions. Non-native species are frequently moved over wide geographic areas, obscuring historical distributional patterns, especially in the absence of accurate introduction records. If these non-native populations become invasive they can cause significant impacts on the environment (Pysek & Richardson 2010), not only affecting biodiversity, but also regional economies (Pimentel *et al.* 2000). Understanding the natal provenances of species is therefore considered a key first step in determining the biogeographic and taxonomic status of species (native vs. non-native), with important implications for management of invasive populations (Sakai *et al.* 2001, Prentis *et al.* 2009). In this respect phylogeography has become an important tool to understand which species are invasive (Schaal *et al.* 2003), determine whether invasive genotypes exists (Lavergne & Molofsky 2007), understand the contemporary evolutionary processes underlying invasiveness (e.g. hybridization), elucidate cryptic invasions (Saltonstall 2002), and infer natal provenances of invasive populations (Le Roux *et al*. 2006, Redfern 2012).

Identifying the natal provenances of invasive taxa has important implications for their effective management, in particular biological control (Le Roux & Wieczorek 2009), for two reasons. First, knowing where invasive populations originated from can direct biological control explorations to regions harbouring suitable host-specific and co-evolved natural enemies (Roderick & Navajas 2003, Goolsby *et al.* 2006). Second, a phylogenetic framework can assist in determining which native taxa are likely to experience non-target impacts (i.e. a phylogenetic centrifugal approach – first testing for non-target impacts on phylogentically closely related taxa, then moving to more distantly related organisms,

sensu Messing & Wright 2006). Many non-native species lack sufficient data on their introduction histories, making molecular approaches valuable for resolving biogeographic uncertainties (e.g. Le Roux *et al.* 2006). Moreover, biogeographic inferences of invasive taxa based on molecular data can also yield unexpected outcomes, e.g. for the true natal structure and diversity of particular taxa (e.g. Ndlovu *et al.* 2013) and primary taxonomy (e.g. Le Roux *et al.* 2006).

An example of how biogeographic uncertainties can hamper on-going management efforts against invasive taxa is that of the peculiar distribution of balloon vine species, genus *Cardiospermum* L. This genus consists of 17 species mostly restricted to the Neotropics, with four species occurring in other parts of the world (Gildenhuys *et al.* 2013). All species found in Africa appear to be native to the Neotropics with the exception of *C. pechuelii*, which is found only in the Namib Desert of Namibia (Gildenhuys *et al.* 2013). Some balloon vine species are considered amongst the worst invasive taxa in many parts of the world, having escaped cultivation following intentional introductions for medicinal and ornamental purposes (Gildenhuys *et al.* 2013). In addition to the extensive movement of balloon vines, the native status of some species is unresolved in many regions, hindering effective management interventions (Simelane *et al.* 2011, Gildenhuys *et al.* 2013). For example, *C. halicacabum* has a near-cosmopolitan distribution, being present on almost every continent and Oceania. The status (native or introduced) of *C. halicacabum* in southern Africa remains unknown (Henderson *et al.* 2001), while it is considered possibly non-native in Australia (Harris *et al.* 2007), North America (Henry & Scott 1981, Bowen *et al.* 2002) and numerous pacific islands (PIER 2013). Similarly, the current distribution of *C. grandiflorum* spans the Americas, Africa and Australasia (Gildenhuys *et al.* 2013), being classified as non-native in Africa (Henderson 2001, Hyde *et al.* 2012a, b), Australia (Carroll *et al.* 2005a), New Zealand (Flora of New Zealand 2012) and some pacific islands (PIER 2013). A third species, *C. corindum*, is also widespread in

the Americas and Africa, and is considered potentially non-native in Africa (Gildenhuys *et al.* 2013).

Invasive balloon vines have significant agricultural and biodiversity impacts (Voll *et al.* 2004, Mc Kay *et al.* 2010, Simelane *et al.* 2011) and therefore various management interventions have been implemented to control invasive populations (Mc Kay *et al.* 2010, Simelane *et al.* 2011, Gildenhuys *et al.* 2013). For example, South Africa's 'Working for Water' programme, a government funded socio-economic initiative aimed at invasive species control, launched a biological control research programme against invasive *C. grandiflorum* in 2003 (Simelane *et al.* 2011). Over the past decade, ten insects and two fungal agents have either been recorded on the target weed in the native South American range or are currently undergoing host-specificity testing in South Africa. Eight of these insects displayed wider host ranges and are capable of feeding and developing on other widespread *Cardiospermum* spp. in South Africa, in particular *C. halicacabum* and *C. corindum* (Mc Kay *et al*. 2010). While the native provenance of South African *C. grandiflorum* is almost certainly Neotropical America (Carroll *et al.* 2005a) the native provenance(s) of *C. halicacabum* and *C. corindum* in South Africa remains speculative (Mc Kay *et al.* 2010, Gildenhuys *et al.* 2013). For instance, the Royal Botanic Gardens, Kew, lists *C. corindum* under its "Flora of Tropical East Africa" collection as native to Pantropical Africa from Sudan, Ethiopia and Somalia to South Africa (Davies & Verdcourt 1998). In the Selmar Schonland Herbarium at Rhodes University in Grahamstown, South Africa, *C. corindum* has even been reported as synonymous to specimens of African *C. pechuelii* (J.J. Le Roux, personal observation). In contrast, an American native range for *C. corindum* is often assumed by others (e.g. Molina-Freaner & Tinoco-Ojanguren 1997, Castellanos *et al.* 1999). Similarly, some authors reported *C. halicacabum* to be indigenous to both the Americas and Africa (Rock, 1972) while others argue in favour of an African native range (Davies & Verdcourt 1998, Hyde & Wursten 2012a, b).

The importance of clarifying the geographic native ranges of all *Cardiospermum* species currently found in South Africa for the implementation of successful biological control of *C. grandiflorum* is evident. If *C. corindum* and *C. halicacabum* are indeed native to southern Africa, only agents that are specific to *C. grandiflorum* can qualify for release in South Africa, and thus far, these agents have proved particularly difficult to rear and test under quarantine conditions (D. Simelane, Agricultural Research Council, personal comm.).

Here, using phylogenetic and phylogeographic approaches, we aim to shed light on the biogeography and status (native or non-native) of selected balloon vine species globally, with an emphasis on southern Africa taxa. We hypothesized that three scenarios can explain the current distribution of these balloon vine species in southern Africa: human-mediated introductions, natural rare long distance dispersal events, and a Gondwanan distribution (Fig. 1). The first scenario would imply an alien status while the latter two scenarios will support a native status.

Scenario	Alien	Native: Long distance dispersal	Native: Gondwana land
			

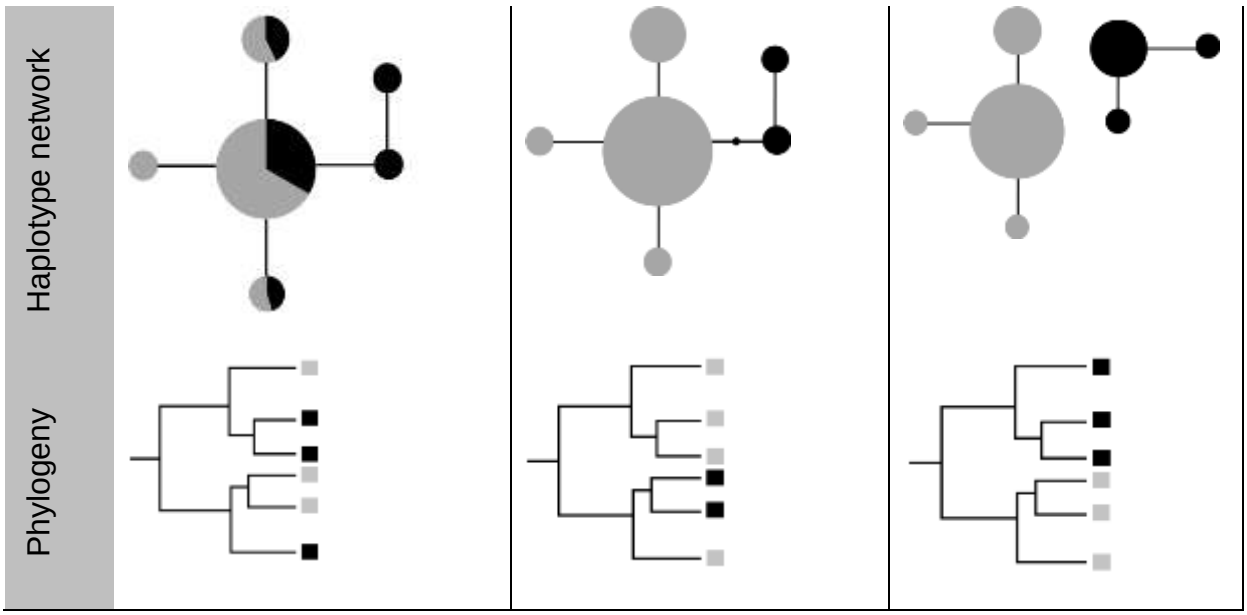


Figure 1: Hypothesised phylogenetic and phylogeographic outcomes for different dispersal/distributional scenarios for *Cardiospermum* species in southern Africa. Hypothetical taxa from Africa are indicated in black and those from South America in grey. A: Human-aided dispersal would lead to an unstructured haplotype network along with a phylogeny indicating polyphyly between South American and southern African accessions due to numerous possible introductions from different sources in the native range as well as shared haplotypes between the native range and southern African sampels since enough evoluitionary time wouldn't have passed for numerous mutations to manifest., B: Rare long distance dispersal would lead to a single group of closely related African haplotypes along with a phylogeny illustrating a monophyletic clade of southern African samples in an overall dominant South American clade and C: Gondwanan distribution would lead to highly structured and diverged haplotypes for South American and southern African samples and a phylogeny with a deep node split and geographic monophyly since time of divergence, ca. 90 MYA, ample time for unique genetic mutations to accumulate.

2.3 Methods

2.3.1 Sampling and DNA extraction

Leaf material was collected from *C. corindum*, *C. halicacabum, C. grandiflorum* and *C. pechuelii* plants from various regions globally (Table S1). Samples were dried and stored on silica gel until further use. Whole genomic DNA was extracted from dried leaf material using the cetyltrimethyl ammonium bromide (CTAB) method as described by Doyle and Doyle (1990), modified by adding 1% PVP- 40T. DNA quality and quantity was assessed using a nanodrop (NanoDrop ND-1000, Inqaba Biotec) and all extractions diluted to a final concentration of ca. 50 ng/µL.

2.3.2 Gene amplification

The *rpl32* chloroplast (cpDNA) gene was amplified using primers rpl32-trnl (5'-CTG CTT CCT AAG AGC AGC GT-3') and rpl32-F (5'-CAG TTC CAA AAA AAC GTA CTT- 3') (Shaw *et al.* 2007). Each 50 µL reaction contained ca. 250 ng of genomic DNA, 200 µM of each dNTP (AB gene, supplied by Southern Cross Biotechnologies, Cape Town, South Africa), 10 pmoles of each primer, 0.2 U Taq DNA polymerase (Kapa Biosystems, supplied by Inqaba Biotech, Cape Town, South Africa), 1 X PCR reaction buffer, and 0.5 mM $MgCl_2$. PCR cycles consisted of initial denaturation of 95 °C for 2 min; 30 cycles at denaturation at 94 °C for 30 s, annealing at 49 °C for 30 s, elongation at 72 °C for 30 s; and final extension at 60 °C for 10 min.

A second cpDNA gene, the *trnL-F* intron was amplified using primers trnL-Fc (5'-CGA AAT CGG TAG ACG CTA CG-3') and trnFf (5'-ATT TGA ACT GGT GAC ACG AG-3') (Taberlet *et al.* 1991). Each 30 µL reaction contained ca. 180 ng genomic DNA, 25 mM of each dNTP (Fermentas Life Sciences, supplied by Inqaba Biotechnology, Pretoria, South Africa), 30 pmoles of each primer, 1 U Taq (Super-therm polymerase, supplied by Separation Scientific, Cape Town, South Africa), 1 µL BSA, 1X PCR reaction buffer, and 2

mM $MgCl_2$. PCR cycles consisted of initial denaturation of 94 °C for 5 min; 30 cycles at denaturation at 94 °C for 60 s, annealing at 55 °C for 60 s, elongation at 72 °C for 90 s; and final extension at 72 °C for 7 min.

The nuclear (nDNA) internal transcribed spacer region (*ITS*), was amplified using the universal primers ITS1 (5'-TCC GTA GGT GAA CCT GCG G-3') and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') and in certain cases ITS5 (5'-GGA AGG AGA AGT CGT AAC AAG G-3') (Fujita *et al.* 2001). Each 30 µL reaction contained ca. 100 ng of genomic DNA, 15 µL Qiagen® multiplex kit (Qiagen, supplied by WhiteHead Scientific, Cape Town, South Africa) and 20 pmoles of each primer. PCR cycles consisted of initial denaturation of 95 °C for 5 min; 40 cycles at denaturation at 94 °C for 30 s, annealing at 54 °C (60 °C for ITS4 and ITS5) for 60 s, elongation at 72 °C for 30 s; and final extension at 72 °C for 10 min. Some sequences yielded ambiguous results (double peaks) and were cloned prior to sequencing. For cloning we used the pGEM-T Easy kit (Promega®, supplied by Anatech, Cape Town, South Africa) with a ligation setup as follows: 2.5 µL Ligation buffer (5X), 0.5 µL vector (pGEM-T Easy) (50ng/ µL), 0.5 µL T4 DNA Ligase and 3 µL Purified PCR Product (varying concentrations between 11-80 ng/ µL). Ligations were done overnight at 4 °C. Transformations were done using Supercompetent DH5-Alpha *E. coli* cells (made in the lab following the modified Hanahan protocol by Tom King; available online: http://parts.igem.org/Help:Protocols/Competent_Cells). Cells were thawed on ice for 5 min and for each transformation 2 µL of ligation product was added to 50 µL of competent cells and incubated on ice for 20 min. Cell mixtures were heat shocked for 45-50 s at 42 °C, followed by incubation on ice for 2 min. 600 µL of liquid luria broth (LB) media was added followed by incubation at 37°C and then shaken for 90 min 200 rpm. Mixtures were plated out on solid LB medium containing ampicillin (100 µg/ml), IPTG (0.1 mM) and Xgal (40 µg/ml). Colonies were grown overnight at 37 °C and colony PCRs done for positive clones using standard cloning primers from Promega®, SP6 (5'-TAT TTA GGT GAC ACT ATA G-

3') and T7 (5' TAA TAC GAC TCA CTA TAG GG-3'). PCR conditions were the same as described above for ITS but with an annealing temperature of 54°C. All PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen, supplied by WhiteHead Scientific, Cape Town, South Africa) and sequenced using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit and an automated ABI PRISM 377XL DNA sequencer (PE Applied Biosystems, Foster City, California, USA).

2.3.3 Phylogenetic analysis

All DNA sequences were aligned using MAFFT (Katoh & Standley 2013) followed by manual inspection and editing in BioEdit v 7.0.5.3 (Hall 1999). Three data matrices were created. A *rpl32* matrix including 134 samples (82 *C. grandiflorum*, 25 *C. halicacabum*, 20 *C. corindum* and 7 *C. pechuelii*), an *ITS* matrix including 116 samples (52 *C. grandiflorum*, 22 *C. halicacabum*, 30 *C. corindum,* 8 *C. pechuelii and* 4 *Paullinia* and *Serjania* outgroup samples) and a combined *ITS* & *trnL-F* dataset including 98 samples (42 *C. grandiflorum*, 23 *C. corindum*, 21 *C. halicacabum,* 8 *C. pechuelii* and 4 *Paullinia* and *Serjania* outgroup samples).

We constructed a haplotype network for the *rpl32* cpDNA dataset, since outgroup data was not available for this gene to be included in the multi gene phylogeny. The haplotype network was built in TCS v 1.21 (Clement *et al.* 2000) using a statistical parsimony network approach with a 95% connection limit.

A Bayesian inference phylogeny was reconstructed for the *ITS* gene using Mr Bayes v 3.2 (Ronquist & Huelsenbeck 2003) that included more samples sampled over a wider geographical range. Akaike information criteria (AIC) scores from jModelTest were used to determine the best fit model (Akaike 1973, Posada 2008). Data was run for 3 million generations sampling every 1000th generation. A consensus tree was build discarding the

first 25% of trees as burn-in and posterior probabilities were calculated using a majority rule consensus method.

A dated Bayesian phylogeny was reconstructed in BEAST v 1.7.5 (Drummond *et al.* 2012) using the combined *ITS* & *trnL-F* dataset. Likelihood tests in Mega v 5.1 (Tamura *et al.* 2013) for rate homogeneity suggested unequal rates for our data, therefore a lognormal relaxed molecular clock was used. To estimate the times of divergence between different taxa/clades we constrained the root of the phylogeny with a normal distribution and mean age of 26.52 ± 3.5 MYA as determined by Buerki *et al.* (2011) for *Cardiospermum*, *Paullinia* and *Serjania*. Three MCMC chains were run for 10 million generations sampling every 1000th generation. Log combiner v 1.7.5 (part of the BEAST v 1.7.5 package) was used to combine the results off all three runs. Results were viewed in Tracer v 1.5 (Rambaut & Drummond 2007) to assess effective samples size values and trees were reconstructed using TreeAnnotater v 1.7.5 (part of beast package), discarding the first 1000 trees. We ran an additional analysis with the same settings as described above, changing the root calibration to an exponential distribution with offset of 21 MYA (5% confidence interval of minimum age from Buerki *et al.* 2013), effectively enforcing a younger tree with younger node ages. Enforcing a younger tree (exponential) increases the chances of node dates to be within the last 10 000 years (effectively human influenced dispersal). Therefore if the dates of the exponential tree predate human influence, we can with reasonable confidence classify the samples as naturally dispersed between continents.

2.4 Results

2.4.1 Sequence variation

The aligned *rpl32* matrix contained 671 base pairs (bp), *ITS* 477 bp and the combined *trnL-F* and *ITS* 855 bp. All DNA sequences have been deposited in GenBank

(accession numbers xxxxxx-xxxxxx – to be submitted upon acceptance) (http://www.ncbi.nlm.nih.gov) (Table S1). The alignment matrix constructed using data generated in this study and additional sequence data obtained from GenBank for outgroups (EU721266.1, EU721327.1, EU721241.1, AY207574.1, EU720494.1, EU720557.1, EU720472.1, AY207571.1) required 8 gaps (indels) all 1 bp in size for *rpl32* (no outgroups) and 45 gaps for *trnL-F-ITS*, ranging from 1 to 12 bp in size. Where gaps longer than 2 bp occurred, the whole dataset was shortened to 1 bp to prevent a single mutational step (inversion, deletion or translocation) being viewed as multiple mutational steps in our network analysis. The best-fit model for the *trnL-F* and *ITS* matrix was identified as the general time reversible model with gamma variable sites (GTR + G).

2.4.2 *Rpl32* haplotype network

Two major disconnected networks were identified, one for *C. grandiflorum* and one including *C. pechuelii*, *C. corindum* and *C. halicacabum* (Fig. 2). Southern African samples of *Cardiospermum corindum* and *C. pechuelii* shared a haplotype and were more closely related to one another than to *C. corindum* samples from South America (Fig. 2). For *C. halicacabum*, specimens from southern Africa (South Africa and Namibia) shared the same haplotype with samples from Mayotte, La Réunion and Hawaii (known non-native ranges) and this haplotype was more closely related to South American samples than other South American haplotypes were to each other (Fig. 2). Unique haplotypes were present for *C. halicacabum* samples from La Réunion, Tanzania, Fiji and Australia, but were still connected to the main South American haplotypes with only a few mutational steps separating them. All *C. grandiflorum* samples from the native and invasive ranges shared the same haplotype, with the exception of samples from Hawaii and Uganda (Fig. 2).

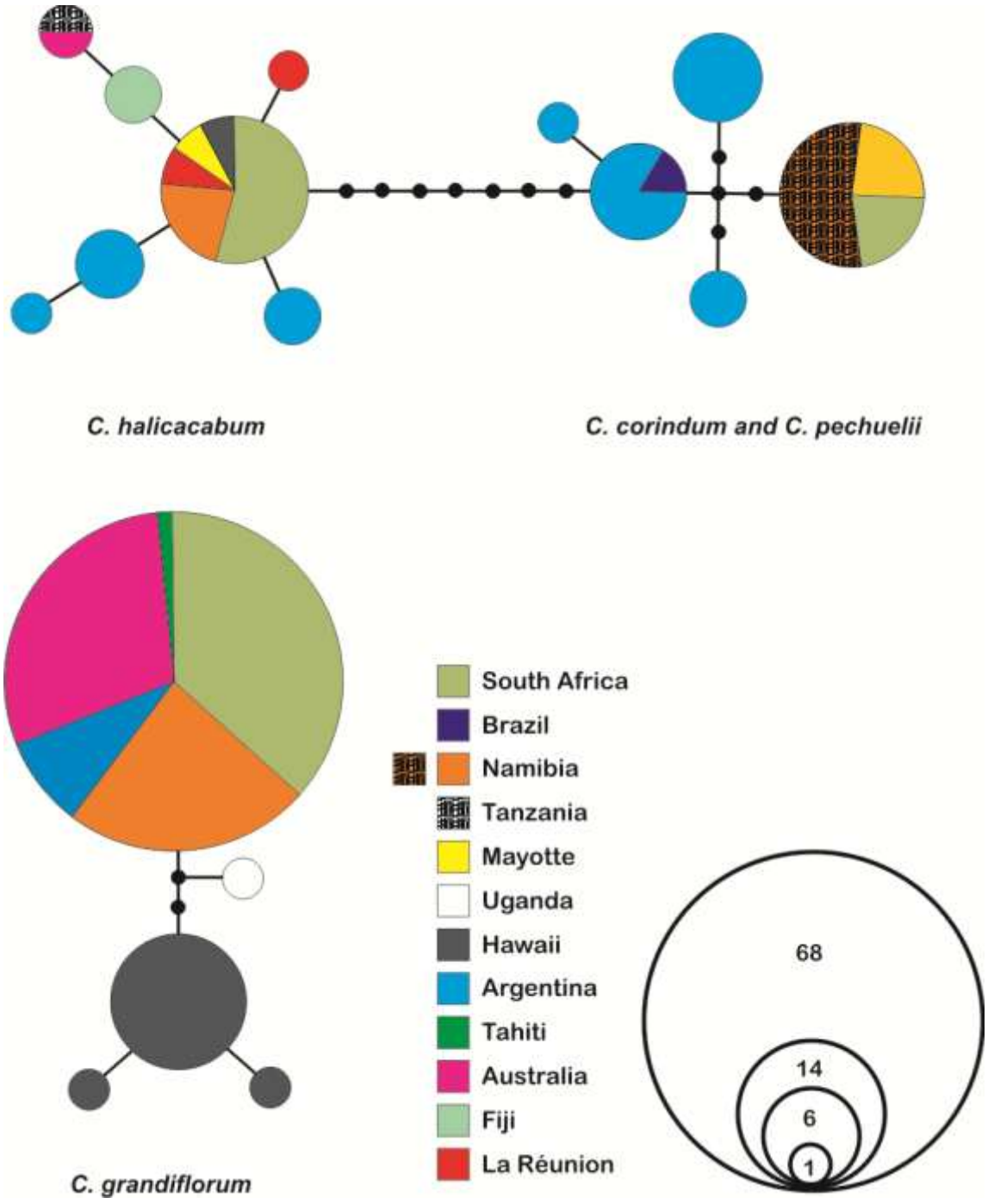


Figure 2: *rpl32* haplotype network of native, invasive and unknown range *Cardiospermum* species. Different countries (or regions e.g. Hawaii) are represented by different colours and *C. pechuelii* haplotypes from Namibia are illustrated by striped patterns. The size of circles indicates frequency of haplotypes and actual haplotype frequencies are indicated in insert.

2.4.3 Phylogenetic analyses

The nDNA *ITS* phylogenetic tree indicated that *C. pechuelii* renders *C. corindum* paraphyletic while the combined nDNA *ITS* and cpDNA *trnL-F* showed that southern African *C. corindum* and *C. pechuelii* are monophyletic at 3.1 MYA, however the support for this southern African clade was relatively low. When enforcing a younger tree, the age of this node still dates at 3.0 MYA (Fig. S1). The overall age of the *C. corindum* clade was dated at 5.7 MYA (Fig. 3 & 4).

Overall, *Cardiospermum halicacabum* formed a well-supported clade (7 MYA) that diverged around 13.3 MYA from the sister clade *C. pechuelii/C. corindum* (Fig. 3). Within the *C. halicacabum* clade polyphyly is evident between samples indicating that the clade was unstructured with respect to geographic regions (Fig. 4).

The phylogeny indicated two well supported distinct clusters within *Cardiospermum grandiflorum*. The first clade contained accessions from Uganda and Hawaii and diverged from the second cluster (Australia, South Africa, Namibia, Argentina, Brazil, and Tahiti) around 5.4 MYA (Fig. 3). The Hawaiian-Uganda cluster further split into country specific clades dated around 4.4 MYA (Fig. 3). The *C. grandiflorum* clade including multiple worldwide samples, showed that global collections were geographically polyphyletic with reference to South American samples with low resolution in this cluster (Fig. 3).

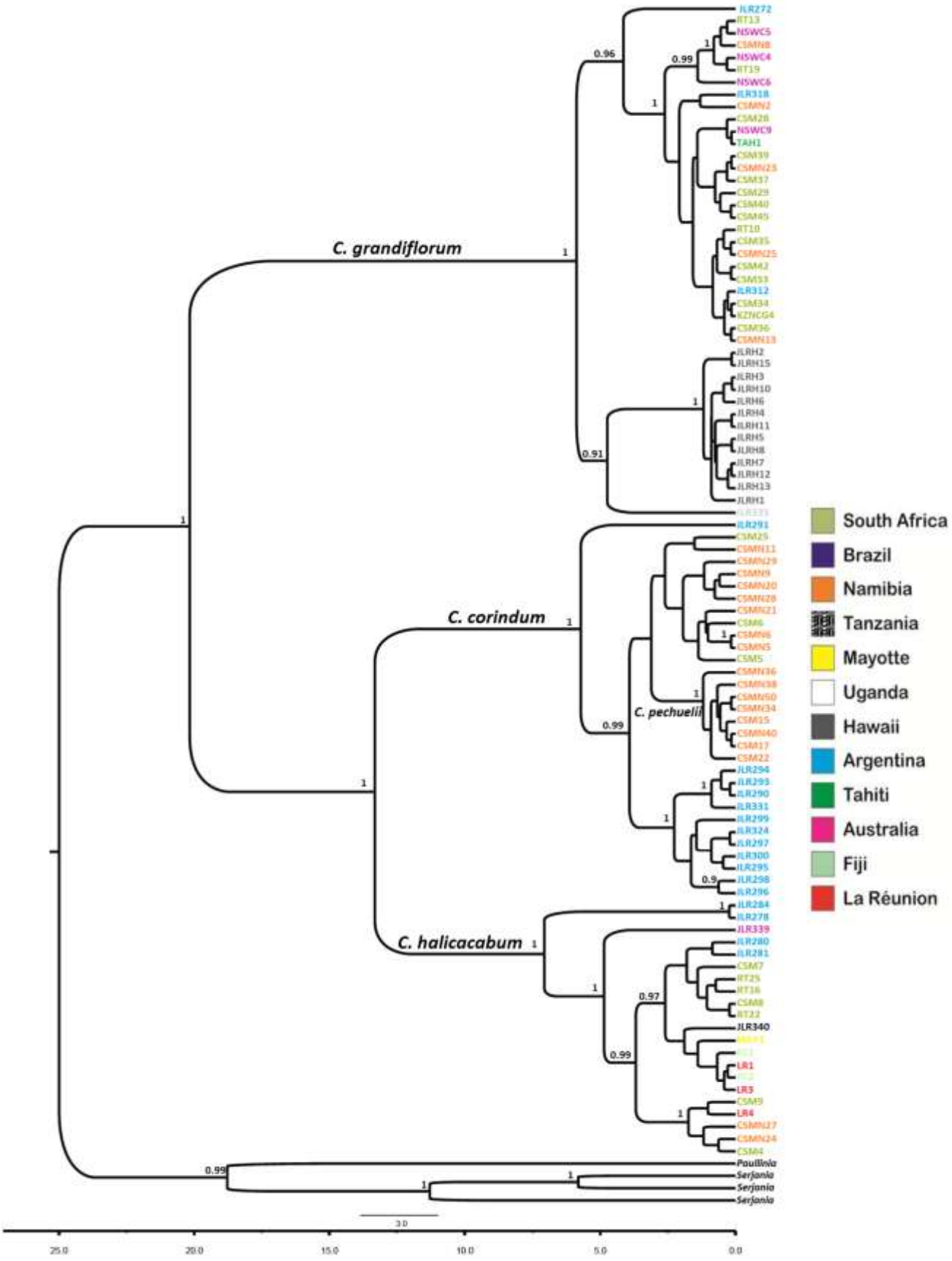

C. grandiflorum
C. corindum
C. pechuelii
C. halicacabum
Paullinia
Serjania
Serjania
Serjania
South Africa
Brazil
Namibia
Tanzania
Mayotte
Uganda
Hawaii
Argentina
Tahiti
Australia
Fiji
La Réunion
25.0
20.0
15.0
10.0
5.0
0.0

Figure 3: BEAST chronogram (based on *ITS* and *trnL-F*) for the genus *Cardiospermum*, including species with unknown native ranges *C. grandiflorum*, *C. corindum*, *C. halicacabum* and the African species *C. pechuelii*. Posterior probabilities higher than 0.9 are indicted at nodes and countries are indicated with different colours, following the key provided in figure 2. The scale bar indicates estimated time since divergence in millions of years (MYA).

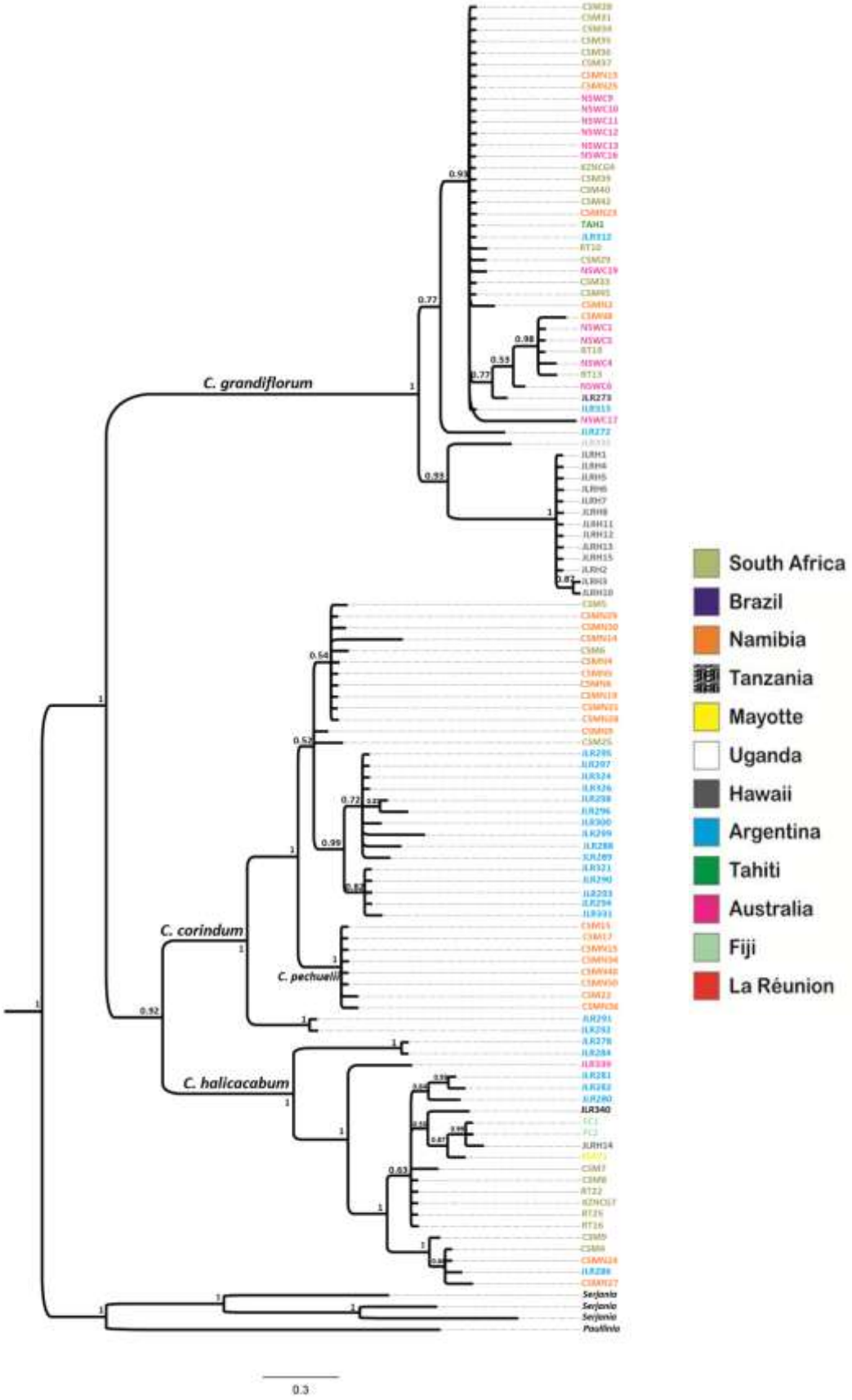


Figure 4: MrBayes cladogram (based on *ITS*) for the genus *Cardiospermum*, including species with unknown native ranges *C. grandiflorum*, *C. corindum*, *C. halicacabum* and the

African species *C. pechuelii*. Countries are indicated with different colours, following the key provided in figure 2. Posterior probabilities are indicated at nodes.

2.5 Discussion

Resolving natal ranges of species with questionable origin is crucial for effective management of alien species as illustrated within the genus *Cardiospermum*. Any questioned native status of a *Cardiospermum* species in Africa due to their presence on the Gondwanan supercontinent is dismissed since the divergence of *Cardiospermum*, *Paullinia* and *Serjania* far post dates continental break-up estimated to have occurred at 90-105 MYA (Richardson *et al.* 2004). However our results strongly suggest that some balloon vine species are native to southern Africa and this has probably occurred through colonization following transoceanic long distance dispersal. This is evident when considering the monophyly for all southern African *C. corindum* and the deep node split dated at 3.9 MYA between South American and southern African *C. corindum*.

Within the *C. corindum* clade all African *C. corindum* accessions exhibited strong geographic monophyly and in the case of the combined tree, all African samples including *C. pechuelii* showed monophyly (Fig. 3 & 4). Two South American accessions (JLR291 and JLR292) were geographically very distant from the other South American *C. corindum* samples (nearest collection site ca. 660 km away, Table S1) and were basal to the core *C. corindum* clade (Fig. 3). The split between these basal accessions and the core clade was estimated at 5.7 MYA. The monophyletic clade shared between all African *C. corindum* and *C. pechuelii* (Fig. 4) supports the hypothesis that African *C. pechuelii* originated from *C. corindum*, following long-distance dispersal. Whether long distance dispersal events occurred more than once remains unresolved due to the different tree topologies for the *ITS* cladogram and combined data dated phylogeny, more accessions with additional markers are needed to fully resolve these relationships and relationship between *C.*

corindum and *C. pechuelii*. In the *ITS* cladogram, the basal position of *C. pechuelii* support multiple independent introductions, while the combined dated phylogeny support a single introduction based on a monophyletic clade for southern African *C. corindum* and *C. pechuelii*. The haplotype network, based on the more conservative *rpl32* cpDNA region, also supported a native origin for southern African *C. corindum* illustrating its close relationship with *C. pechuelii* (Fig. 2). While both *C. corindum* and *C. pechuelii* occur in Namibia they occupy very distinct habitats with *C. corindum* only occurring in moist subtropical areas, e.g. of the Waterberg Plateau, as opposed to *C. pechuelii* that is restricted to arid areas of the Namib Desert. This habitat preference is also evident in their strikingly different morphologies with *C. corindum* clearly adapted for the sub-tropical environments as opposed to the water efficient morphology of the arid adapted *C. pechuelii*. The phylogenetic relationships retrieved here between *C. pechuelii* and *C. corindum* make these species a promising system to investigate rates of adaptation and evolution of reproductive isolation at the population level.

Transoceanic dispersal and exchanges between the Americas and Africa are not uncommon. For example, it is estimated that *Maschalocephalus dinklagei* (family Rapateaceae) and *Pitcairnia feliciana* (family Bromeliaceae) dispersed from South America to West Africa around 6-8 MYA (Givnish *et al.* 2004). Similarly Bartish *et al.* (2010) concluded that long distance dispersal is the main mechanism for pantropical range expansion for the subfamily Chrysophylloideae between Africa, South America, New Caledonia and Australia. Renner (2004) noted that 110 angiosperm genera in 53 families have representatives in both South America and Africa, and even though molecular phylogenies were only available for a limited number of these taxa, they showed that wind- or water-mediated long distance dispersal has been implicated for most of these species across the tropical Atlantic. Recent long-distance dispersal events also explain the cosmopolitan distributions of other species in the balloon vine family (Sapindaceae), such

as *Dodonea viscosa* (Harrington & Gadek 2009). Within the past 2 million years, this species group has colonized numerous regions globally and is currently considered native in many countries in six continents (Harrington & Gadek 2009).

For *C. halicacabum*, native range biogeographic patterns were less clear than for *C. corindum*, but the phylogeny suggested an alien status in southern Africa. In the phylogeny polyphyly is evident showing that South American and southern African *C. halicacabum* samples are not geographically structured (Fig. 3). However a well-supported clade including samples from Mayotte, Fiji, Hawaii and Tanzania, but excluding South American samples, make it difficult to determine the native or alien status in these countries. The cpDNA network and multi gene phylogeny do not provide conclusive evidence one way or the other for these accessions. Our phylogeny also points towards a native status for *C. halicacabum* in Australia given the deep node split of 4.8 MYA following Cowie and Finlayson (1986) as opposed to a non-native status as previously suggested by Harris *et al.* (2007). However, in order to completely shed light on the native or alien status of *C. halicacabum* in Australia and other regions in question like Mayotte, Fiji, Tanzania and Hawaii more collections over a wider geographical area have to be included for both the countries in question and native South America.

The alien status of *C. grandiflorum* in southern Africa, Australia and other parts of the world is supported with polyphyly for samples worldwide (Fig. 3) and a single haplotype shared by all southern Africa, South American, Tahiti and Australian samples (Fig. 1). However, our results call into question a contemporary South American origin for tropical Uganda and Hawaiian populations. We recommend reviewing the species' status in these areas. These samples are several mutational steps from all other *C. grandiflorum* samples in the haplotype network. Also the phylogenetic tree shows the tropical Africa *C. grandiflorum* sample and Hawaiian samples splitting from the closest related South

American samples at 5.8 MYA, suggesting *C. grandiflorum* is possibly native in tropical Africa. This taxon has been variously regarded as native (Perreira *et al.* 2012) and alien (Mosango *et al.* 2001) in Uganda, and indeed the fruit morphology differs substantially between Ugandan specimens and those found in South Africa and the Neotropics (S.P. Carroll, personal observation). In Hawaii *C. grandiflorum* is regarded as an invasive weed (PIER 2013), however the split from the rest of the samples at 5.8 MYA, long predates human influence. In Hawaii *C. grandiflorum* also displays little evidence of aggressive invasiveness (J.J. Le Roux, personal observation). Even when enforcing the youngest possible age for this split (Fig. S1), it dates to a 95% minimum age no younger than 5.4 MYA. The deep monophyly of the African and Hawaiian samples was unexpected and potentially questions the non-native status of *C. grandiflorum* in these areas. Alternatively this pattern could result from inadequate sampling of the South American source pool, or through introductions to Hawaii from tropical African regions. Further sampling is required to distinguish these possibilities.

2.5.1 Long distance dispersal within *Cardiospermum*

While there are many ways by which plants can travel long distances, retrospective inferences will remain speculative as it is nearly impossible to determine specific mechanisms with complete accuracy (Guillespie *et al.* 2012). However the morphology of a taxon may give insights into likely dispersal strategies. For example, some species are well-adapted for wind dispersal such as various *Miscanthus* species, *Calluna vulgaris* and *Erica cinerea* (Quinn *et al.* 2011, Soons & Bullock 2008) while others disperse by animals, mostly frugivorous birds (Nogales *et al.* 2012). Other species, lacking morphological adaptations for wind or water dispersal, e.g. *Nothofagus* from Australia to New Zealand, have likely rafted long distances (Knapp *et al.* 2005).

Cardiospermum fruit are well-adapted for water dispersal, making ocean dispersal using currents far more feasible than any of the above mentioned dispersal strategies. The

fact that *Cardiospermum grandiflorum* seeds retain viability after fruit were floated for more than 6 months provides evidence supporting the ability of *Cardiospermum* fruit to travel long distances in ocean waters (Gildenhuys, unpubl. data). Carroll *et al.* (2005) also noted the ability of *C. grandiflorum* to spread between landmasses during storm events, such as the balloon vine invasion in Rarotonga's national parks following a tropical cyclone. All *Cardiospermum* dispersal events are estimated to be geographically relatively young and ocean currents are unlikely to have changed within the time frame of intercontinental movements in *Cardiospermum* (i.e. 3-7 MYA). An example of another species in the Sapindaceae family that likely utilised ocean currents to become a widespread native species is *Dodonaea viscosa* (Harrington & Gadek 2009). Floating experiments for this species found that 30% of fruit floated after 100 days and 79% of seeds germinated after being soaked in sea water for 6 months (West 1980).

2.6 Conclusions

While some of the past biogeographic uncertainties among balloon vines remain, our work strongly suggests that *C. pechuelii* and *C. corindum* are native in Africa, but that *C. grandiflorum* and *C. halicacabum* in southern Africa are of recent South American origin. A very important implication of our research is that biological agents currently identified for the control of *C. grandiflorum* should not be released in South Africa without fuller understanding of possible non-target effects on native *C. corindum* and *C. pechuelii.* In contrast, possible non-target impacts on introduced Neotropical *C. halicacabum* are a lesser concern from the standpoint of native species protection. Biological control of non-native *Cardiospermum* may be pursued by continued development of more species-specific Neotropical agents, or by aiding adaptation in native South African natural enemies of *Cardiospermum* species to better exploit the invasive members of the genus (*sensu* Carroll 2011).

2.7 Acknowledgements

We thank David Simelane and Andries Fourie who kindly provided us with *Cardiospermum* DNA samples from South America and Kate Le Roux, Hannes Wiese, Mandelise van Loggerenberg and Ralph Riley for field assistance. We are grateful for sampling permits from NSW Department of Environment and Conservation and Queensland Government Department of Environment and Resource Management. We thank the Australian Tropical Herbarium and Australian National Herbarium for herbarium specimens. Sven Buerki provided the root age and confidence intervals for calibration of the *Cardiospermum*, *Serjania* and *Paullinia*. Financial support was provided by the DST-NRF Centre of Excellence for Invasion Biology and the Working for Water Programme through their collaborative project on “Research for Integrated Management of Invasive Alien Species”. E.G. acknowledges the South African National Research Foundation’s (NRF) Scarce Skills scholarship programme. J.L.R. also acknowledges Stellenbosch University’s Sub-committee B “Young Researchers Fund” and the NRF Thuthuka Programme for research funding. S.C. acknowledges support from the School of Life Sciences at the University of Queensland, St. Lucia.

3. From the Neotropics to the desert: Evidence for rapid diversification of a plant following extreme long distance dispersal.

Intended journal for submission: *Annals of Botany*

3.1 Abstract

Background and Aims: Extreme long distance dispersal by plants is more frequent than previously thought and may lead to rapid diversification due to strong founder effects and strong selection in novel environments. Such an example may be present in the genus *Cardiospermum* evident by a close relationship between the Namib Desert endemic *Cardiospermum pechuelii* and subtropical *C. corindum* previously observed. These species occupy disjunct and distinct habitats, with morphological adaptations specific for their environments. We aim to investigate the rapid divergence of *C. pechuelii* following long distance dispersal of *C. corindum* from South America to southern Africa, specifically to determine whether the former evolved from the latter.

Methods: To resolve the phylogenetic relationships between these two species we used DNA sequencing data from three genetic markers (cpDNA *rpl32* and *trnL-F* and nDNA *ITS*) to reconstruct evolutionary relationships among these taxa using tree building and network analyses. We genotyped accessions using amplified length polymorphism (AFLP) markers for further genetic structure and diversity analyses between sampled populations.

Key results: Our haplotype network (cpDNA only) indicated that southern African *C. corindum* and *C. pechuelii* are more closely related to each other than to South American *C. corindum* and is supported by our phylogeny, grouping all southern African samples into a single monophyletic clade. Bayesian assignment analyses of AFLP data supported our phylogeny, indicating that *C. corindum* samples from southern Africa share the majority of their genetic make-up with *C. pechuelii*, and that these two species appear distinct from *C. corindum* from South America. AFLP diversity was the highest for *C. corindum* populations from South America, while *C. pechuelii* showed the lowest diversity.

Conclusions: We confirm the long distance dispersal of *C. corindum* from South America to southern Africa as previously suggested and show the consequent evolution of the

desert species *C. pechuelii* from South American *C. corindum.* We conclude that further morphological and additional genetic analyses are needed to unravel the specific evolutionary paths between southern African *C. corindum* and *C. pechuelii*. The complex relationship between these species makes this study system interesting to test various species concepts.

Keywords: *Cardiospermum*, endemic, evolution, extreme long distance dispersal, Namib Desert, speciation, species concepts.

3.2 Introduction

Extreme long distance dispersal is more prevalent in plants than previously thought, with numerous examples of even intercontinental exchanges (Givnish *et al*. 2004, Renner 2004, Tremetsberger *et al*. 2005, Pannell & Dorken 2006, Nettel & Dodd 2007, Bartish *et al.* 2010, Takayama *et al*. 2013). Such long distance dispersal events are often characterised by reduced genetic variation following initial establishment i.e. strong founder effects (Austerlitz *et al*. 1997). It is therefore conceivable that such strong founder effects along with novel environmental conditions may lead to rapid diversification (or extinction) due to strong genetic drift and selection (Carson & Templeton 1984, Templeton *et al*. 2008). Most rapid speciation events following a single colonization have been documented in insular habitats (Baldwin & Sanderson 1998, Carlquist *et al*. 2003, Wagner & Funk 1995, Price & Wagner 2004, Knope *et al*. 2012). Here we investigate the possibility of rapid diversification following long-distance dispersal from South America to southern Africa within the balloon vine genus *Cardiospermum* L. (Sapindaceae).

Cardiospermum (17 species), commonly known as balloon vines, are mostly restricted to the Neotropics with a few exceptions (Gildenhuys *et al*. 2013). Numerous species have also been introduced globally to areas where they are now considered invasive (Chapter 2). For example, in southern Africa *C. grandiflorum* and *C. halicacabum* are considered to have been introduced from South America and invasive, while *C. corindum* is regarded as native to both southern Africa and South America (Chapter 2). A phylogenetic treatment of the group indicated that the latter likely represents an extreme long distance dispersal event from South America to southern Africa, around 3 MYA (Chapter 2).

In Namibia *C. corindum* is found in tree and shrub savannah biomes (Fig. 1C), different from the mostly tropical areas in South America. More surprising is that the

African endemic balloon vine species, *C. pechuelii*, shares a monophyletic relationship with southern African *C. corindum* (while rendering *C. corindum* paraphyletic, Chapter 2). *Cardiospermum pechuelii* is an arid adapted shrub endemic to the rocky outcrops of the Namib Desert (Fig. 1F) of Namibia (Chapter 2). Despite their close phylogenetic relationship, southern African *C. corindum* and *C. pechuelii* have notably different morphological traits and occupy very distinct and disjunct distributions (Fig. 1A, B, E & F). The climate in areas of Namibia where *C. corindum* is found, such as the Waterberg Plateau region, and the arid areas where *C. pechuelii* is found, such as the Namib Desert, vary greatly and may explain their distinct morphologies and habits. The Namib Desert is the world's oldest desert and harsh environmental conditions have resulted in flora characterised by xerophytic taxa such as the endemic *Welwitschia mirabilis, Acanthosicyos horrida* and *Arthraerua leubnitziae* (Van Damme 1991). The Namib Desert is characterised by little rain with coastal fog and dew being the primary water source and, along with limited soils and the dominance of bedrock surfaces, represents a harsh environment to plants (Van Damme 1991). *Cardiospermum pechuelii* have smaller leaf surface areas, a more stunted shrub-like growth form, and reduced fruit capsules, compared to *C. corindum*, potentially adaptations to compensate for water loss (Fig. 1B & C). When grown alongside *C. corindum* form South America and southern Africa in a greenhouse, *Cardiospermum pechuelii* retained its morphological differences, ruling out phenotypic plasticity in explaining morphological differences (David Simelane, personal comm.). The strikingly different morphologies between *C. pechuelii* and *C. corindum* make the close phylogenetic relationship between the two species peculiar and an interesting case to investigate how quickly adaptation and diversification can occur.

Expanding on the previous study of the *corindum-pechuelii* complex (Chapter 2), we here aim to resolve the historic relationship between *C. pechuelii* and *C. corindum* using a

phylogenetic and population genetics approach to determine the relatedness between *C. pechuelii* and *C. corindum* occurring in southern Africa and South America.

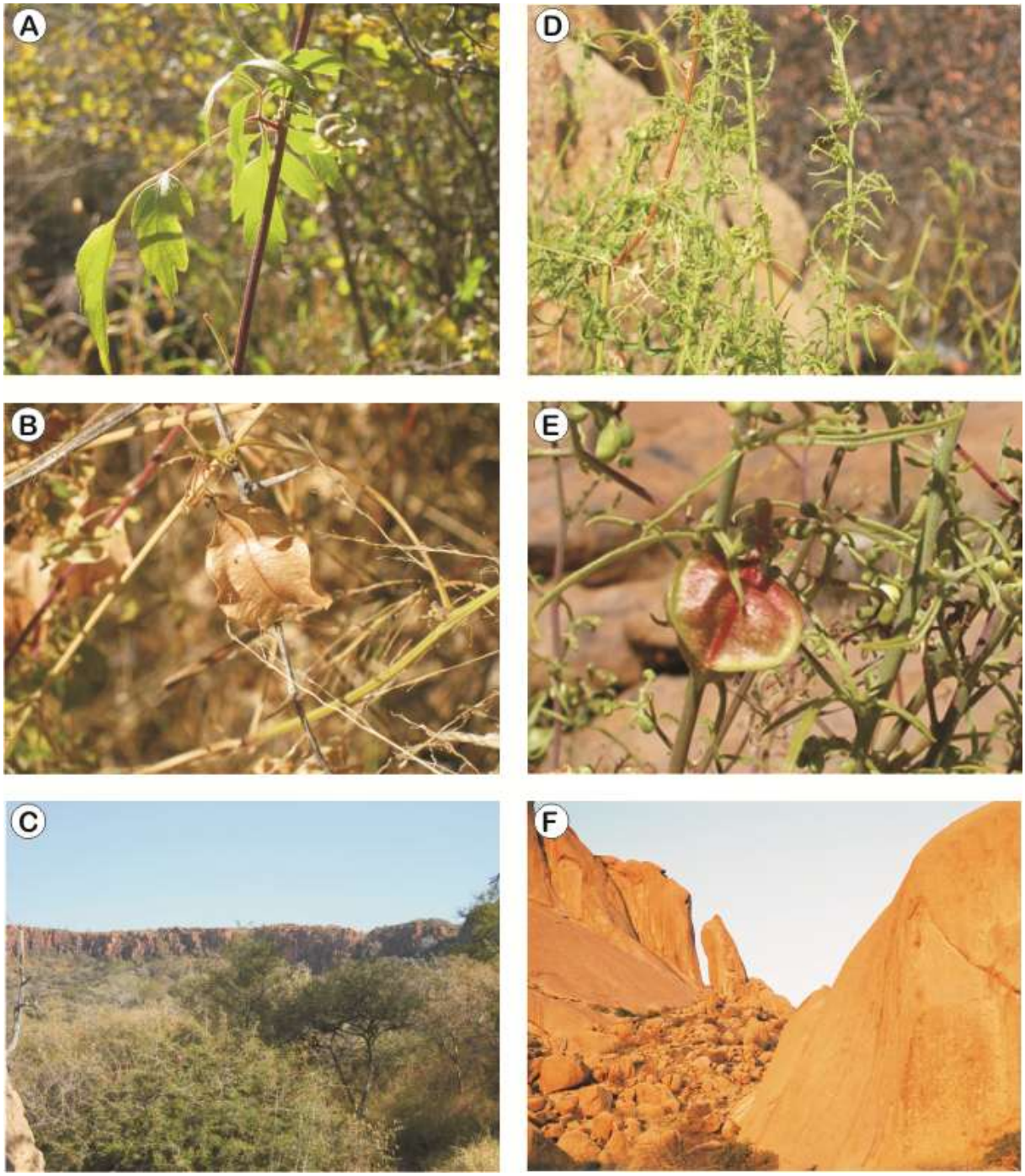


Figure 1: Typical morphological and environmental characteristics of *Cardiospermum corindum* (A-C) and *C. pechuelii* (D-F) in their natal ranges in Namibia: leaf morphology (A and D); fruit morphology (B and E); and habitats occupied: the savannah Waterberg region (C) and rocky outcrops of the Namib Desert (F).

3.3 Materials and methods

Leaf material was collected for 15 *C. corindum* plants from South America and 9 for southern Africa and for 32 *C. pechuelii* from Namibia (Fig. 2) (Table S1). Samples were dried and stored on silica gel until further use. Whole genomic DNA was extracted from dried leaf material using the cetyltrimethyl ammonium bromide (CTAB) method as described by Doyle and Doyle (1990), modified by adding 1% PVP- 40T. DNA quality and quantity was measured using a nanodrop (NanoDrop ND-1000, Inqaba Biotec) and diluted to a final concentration of ca. 50 ng/µL.

3.3.1 Gene and AFLP amplification

Three gene regions, two chloroplast rpl32 and trnL-F and one nuclear, internal transcribed spacer region (*ITS*), were amplified using the methods described in chapter 2. PCR products for the ITS region were cloned due to sequence ambiguity following the methods described in chapter 2. PCR amplification of AFLP fragments was done using the 'universal' protocol described by Blignaut *et al*. (2013). Briefly, following digestion and pre-selective PCR, selective PCR amplification with three different fluorescently labelled EcoRI primers was done. Only one primer pair (EcoR1-CAT and Mse1-CTT) resulted in good quality and repeatable profiles (see results) and was retained for further analyses. In total ninety-nine samples were amplified with AFLP markers of which fifty-three were duplicates to estimate error rates and repeatability of all AFLP loci.

3.3.2 Phylogenetic analysis

DNA sequences were aligned using MAFFT (Katoh & Standley 2013) followed by manual inspection and editing in BioEdit v 7.0.5.3 (Hall 1999). Two data matrices were created, a chloroplast matrix including 45 samples and one including all three gene regions for 39 samples.

A cpDNA haplotype network was constructed from the *rpl32* and *trnL*-F dataset using a statistical parsimony network approach with 95% connection limit in TCS v 1.21 (Clement *et al*. 2000). A Bayesian inference phylogeny was reconstructed for all three gene regions using Mr Bayes v 3.2 (Ronquist & Huelsenbeck 2003). jModelTest (Posada 2008) and the Akaike information criterion (Akaike, 1973) was used to determine the best fit model for our data. The Bayesian model was run for 2 million generations sampling every 1000th generation. A consensus tree building method was used, discarding the first 25% of trees as burn-in. Posterior probabilities (PP) were calculated using a majority rule consensus method to access tree topology support.

3.3.3 AFLP genotyping

Genemarker Version 2.2.0 (SoftGenetics, LLC, CA, USA) was used to visually investigate and remove failed AFLP samples. Peak scoring followed the methods proposed by Ley and Hardy (2013). Briefly, after failed samples were removed, data was imported into PEAKSCANNER© software v 1.0 (Applied Biosystems) where peaks were automatically scored using default settings. Then the sizing table exported from PEAKSCANNER© was imported into tinyFLP v 1.22 (Arthofer 2010) to optimize peak selection. Settings for tinyFLP followed those described by Ley and Hardy (2013) with the following adjustments: minimum peak height, 1000; minimum size, 80bp; maximum size, 400 bp; and maximum frequency, 100%. The modified dataset was transformed into a binary matrix using tinyCAT (Arthofer 2010). Duplicate samples were imported into a new file to estimate broad sense heritability (H^2) of each loci in SPAGeDi v 1.4 (Hardy & Vekemans 2002) using 1000 permutations and coding data as haploid organisms (H^2 neglects dominance effects). The F_{ST} output for each loci from the duplicate dataset was used as estimates of H^2 of each loci (Ley & Hardy 2013), after which any loci with a H^2 less than 0.7 ($P < 0.05$) were dropped for further analysis.

A new trimmed matrix was constructed for further analyses excluding all duplicate samples and loci with H^2 lower than 0.7. To assess the genetic structure the matrix was imported into Structure v 2.3.4 (Falush *et al*. 2007) and individual genotypes were assigned to genetic clusters using a Bayesian approach. Simulations were run for 1 to 7 genetic clusters (*K*) following the admixture model and correlated allele frequencies. Ten runs of 500 000 iterations and burn-in of 20 000 were run for each *K* value. Structure harvester (Earl & von Holdt 2012) was used to determine the optimal number of *K*, using the Δ K method of Evanno *et al*. (2005).

All genetic diversity analyses were done in GenAlEx6 v 6.50 (Peakall & Smouse 2006). Population statistics included analysis of molecular variance (AMOVA), number of effective alleles (*Ne*), Nei's gene diversity (*h*) and Shannon's diversity index (*I*). Pairwise population structure was tested using Nei genetic distance. To visualise genetic clustering a neighbor-net network was constructed based on genetic distances.

3.4 Results

3.4.1 Sequence variation

The aligned combined *rpl32* and *trnL-F* matrix contained 1460 base pairs (bp) and the matrix containing all three genes 2046 bp. The *rpl32* and *trnL-F* matrix (no outgroups) required 27 gaps (indels) for alignment, ranging from 1 to 6 bp in size. Gaps longer than 2 bp, were shortened to 1 bp to prevent a single mutation (inversion, deletion or translocation) to be viewed as multiple mutations. The combined alignment matrix constructed using data generated in this study and two additional outgroup sequences obtained from chapter 2 required 29 gaps.

3.4.2 Haplotype network

The haplotype network illustrated a close relationship between samples from southern African *C. corindum* and *C. pechuelii*, even sharing two haplotypes (Fig. 3). Moreover, *C. corindum* from southern Africa and *C. pechuelii* were more closely related to each other than to any *C. corindum* samples from South America, with the latter isolated by several mutational steps (Fig. 3).

3.4.3 Phylogenetic tree

The best-fit model for the matrix containing all three genes was identified as the general time reversible model with gamma variable sites (GTR+G). *Cardiospermum corindum* samples from South America fell basal to southern African *C. corindum* and *C. pechuelii* samples (Fig. 3). One South American sample geographically isolated from other South American samples (JLR291) did not fall into the overall South American clade, but was basal to this clade. Therefore southern African accessions rendered South American *C. corindum* paraphyletic. A well supported mixed and monophyletic southern African clade (PP = 1) included *C. pechuelii* and *C. corindum*, further structure in this clade was weakly supported (Fig. 3). However, all *C. corindum* render *C. pechuelii* paraphyletic.

3.4.4 Genetic diversity and structure analysis

A total of 41 loci for 46 samples were considered highly reproducible with H^2 scores of higher than 0.7. Structure analysis identified two genetic clusters (Fig. 2, Fig S1). The Bayesian clustering method showed that all of *C. pechuelii* and *C. corindum* samples from southern Africa share the same genetic make-up and form one cluster (Fig. 2).

AMOVA results showed that 37% of the AFLP genetic variation resided within populations and that 63% of the variation resided among populations (Table 1A). Even though sample sizes were higher for *C. pechuelii* than for *C. corindum*, all genetic diversity results (number of effective alleles, Shannon's information index, expected heterozygosity

and % polymorphic loci) indicated the highest level of diversity for *C. corindum* from South America and lowest for *C. pechuelii* populations, except for % polymorphic loci which was lowest for southern African *C. corindum* (Table 1B). Nei pairwise genetic distances not only showed that genetic structure is higher between South American and southern African *C. corindum* than between southern African *C. corindum* and *C. pechuelii*, but also that differentiation is lower between *C. pechuelii* and South American *C. corindum* than between *C. corindum* from southern Africa and South America (Table 1C).

In agreement with our STRUCTURE results, the neighbor-net network based on genetic distances (Hamming distances) estimated from AFLP data with high bootstrap support (>70%) clustered all *C. corindum* southern African samples and *C. pechuelii* samples together (Fig. 4) with the South American *C. corindum* forming a distinct isolated cluster.

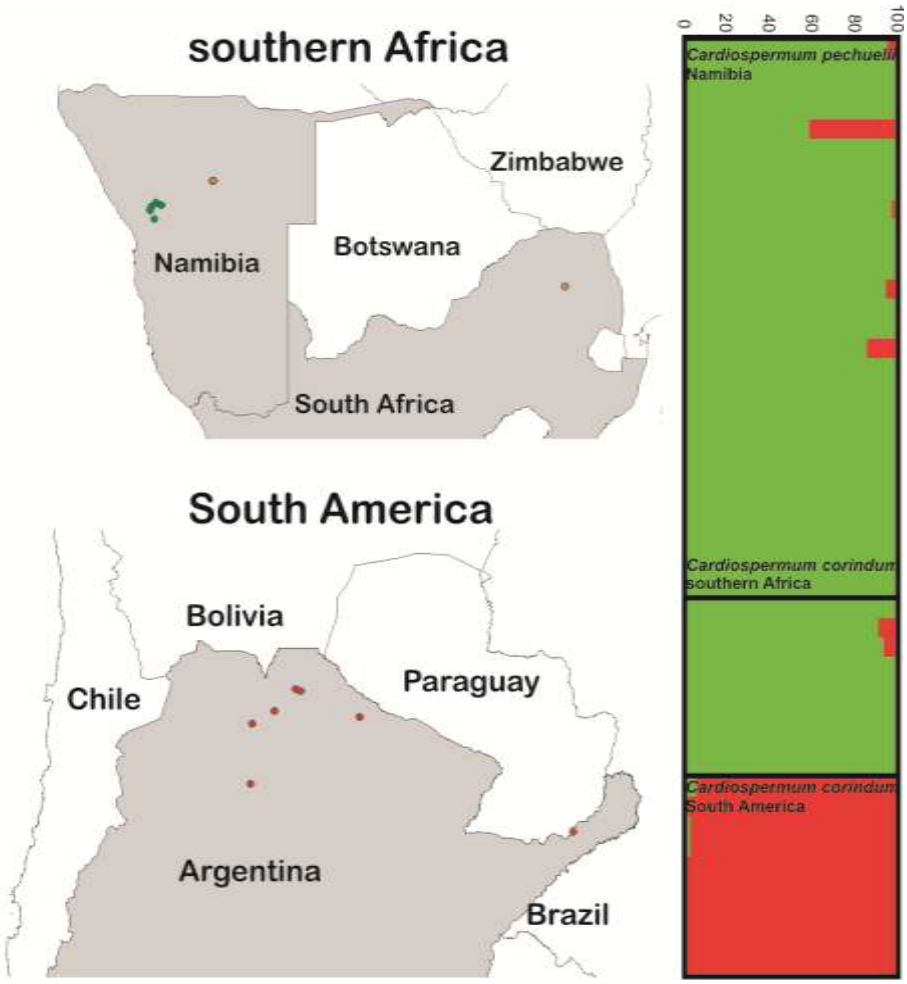


Figure 2: Sampled localities of populations of *C. corindum* in southern Africa (orange) and *C. pechuelii* (green) in Namibia and *C. corindum* in South America (red) (left). Population genetic structure was estimated from individuals from these populations using Bayesian assignment tests. For *K*=2 (number of genetic clusters) each horizontal bar illustrates a different individual and each colour the proportion of its genome assigned to each genetic cluster (right).

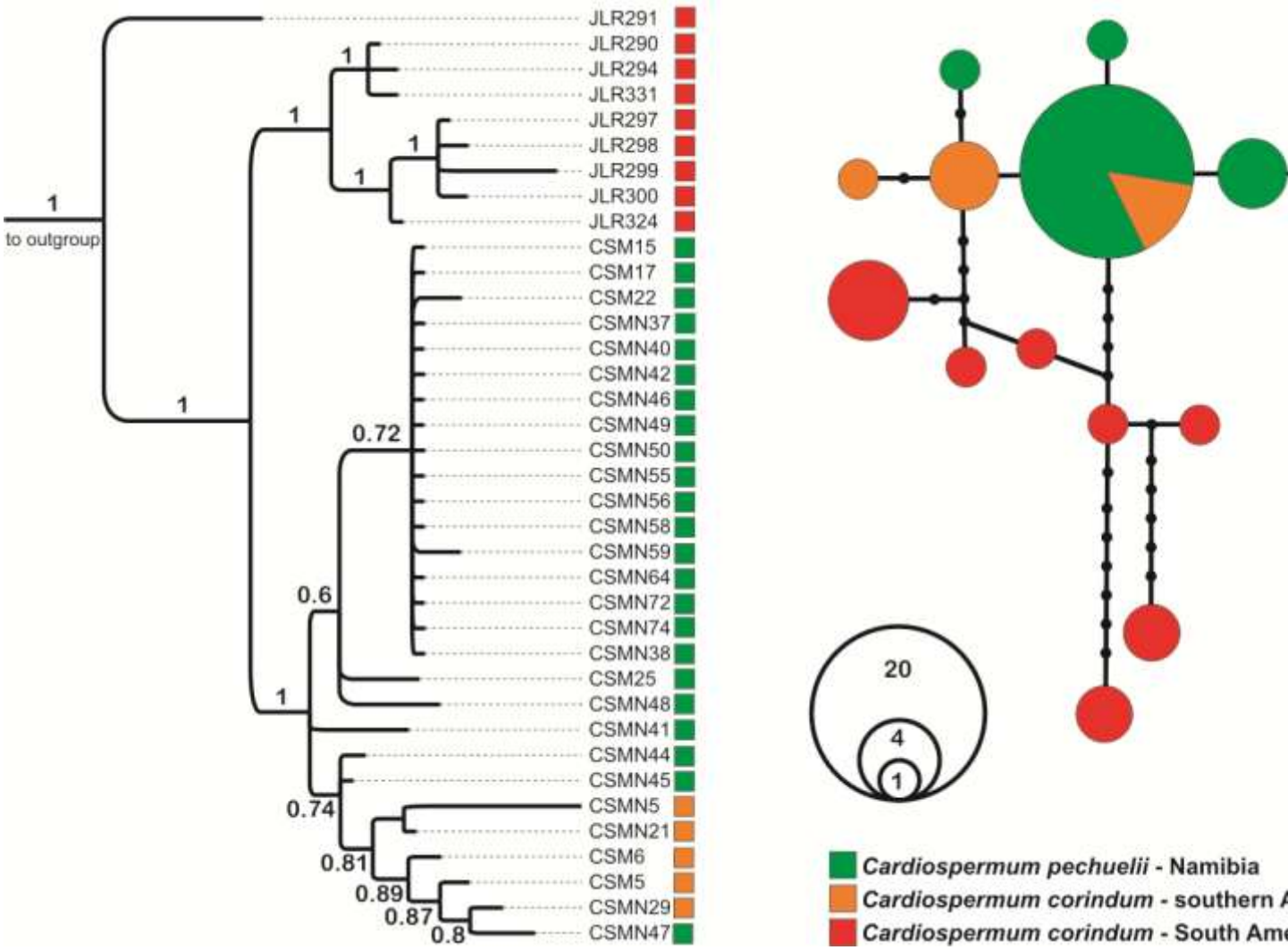


Figure 3: Bayesian phylogeny based on two cpDNA (*rpl32* and *trnL-F*) and one nDNA (*ITS*) gene sequences indicating the relationships among *Cardiospermum corindum* from South America and southern Africa and *C. pechuelii* from Namibia (left hand side). Branch support is given as posterior probabilities and accessions colour coded according to species and geographic affiliations. A haplotype network (right hand side) shows relationships for these species based only on the cpDNA. The sizes of the circles in the network are proportional to the amount of individuals sharing each haplotype and actual haplotype frequencies are indicated in insert.

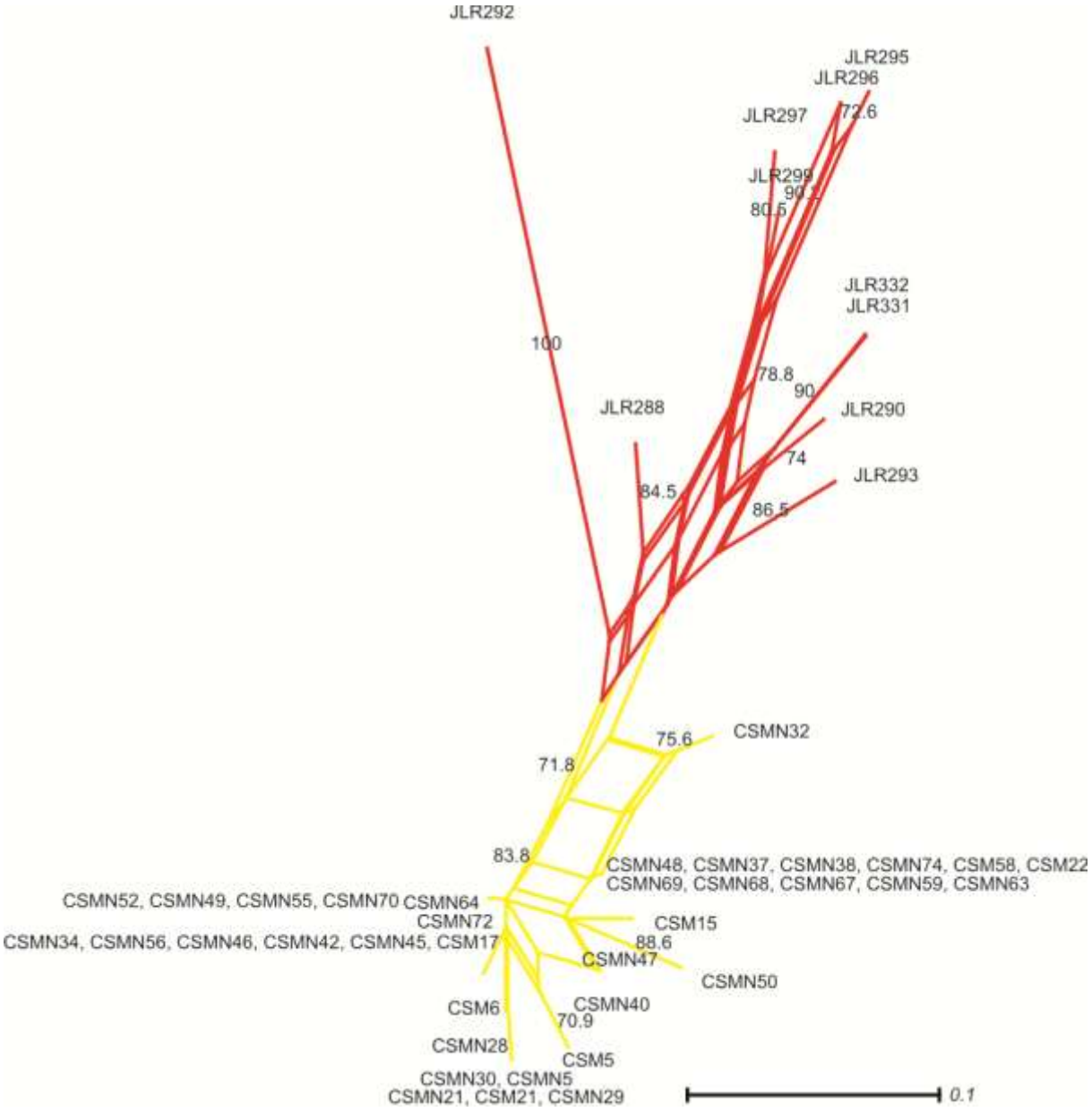


Figure 4: Neighbor-net network based on genetic distances estimated from AFLP data for *Cardiospermum corindum* from South America (red) and *C. pechuelii* and *C. corindum* from southern Africa (yellow). Only bootstrap support >70% are shown in the network.

Table 1: Genetic diversity results based on AFLP analyses for *Cardiospermum corindum* from southern Africa and South America and *C. pechuelii* from Namibia. A) AMOVA results indicating the distribution of genetic variation within and among populations. B) Genetic diversity analyses: effective alleles (*Ne*), Shannon's information index (*I*), expected

heterozygosity (*He*) and % polymorphic Loci (*p*). C) Pairwise Population Matrix of Nei Genetic Distance.

A

	SS	MS	Est. Var.	%
Among Pops	70.2	35.1	2.6	63%
Within Pops	65.1	1.5	1.5	37%

B

	Ne	*I*	*He*	*p*
C. pechuelii	1.030	0.035	0.019	21.95%
C. corindum southern Africa	1.074	0.061	0.041	12.20%
C. corindum South America	1.290	0.285	0.182	65.85%

C

	C. pechuelii	*C. corindum* southern Africa	*C. corindum* South America
C. pechuelii	0.000		
C. corindum southern Africa	0.050	0.000	
C. corindum South America	0.141	0.153	0.000

3.5 Discussion

Here we show how extreme long distance (inter-continental) dispersal can quickly lead to diversification following contact with novel environmental conditions. Endemic *Cardiospermum pechuelii* evolved after *C. corindum* dispersed from South America to southern Africa due to the harsh Namib Desert conditions.

Our results support previous reports that *C. corindum* from southern Africa is likely the result of long distance dispersal (Chapter 2) and we show that this event likely only occurred once (or possibly multiple dispersal events from the same source population). Natural long distance dispersal (historic) as opposed to anthropogenic movements (recent events) is supported by our haplotype network indicating that southern African samples are several mutational steps away from South American samples and no mixed haplotypes between South America and southern Africa are present (Fig. 3). Also a well-supported single monophyletic clade for all southern African samples was identified in our phylogenetic analysis (Fig. 3). As discussed in chapter 2 one would expect an unstructured haplotype network and polyphyletic phylogenetic relationships between southern African and South American taxa if human-mediated introductions explained *C. corindum*' s presence in southern Africa. Further support for natural long distance dispersal comes from the reduced genetic diversity (Table 1B) observed for southern African samples compared to South American samples in our AFLP data. Anthropogenic movements often lead to equally high or higher [e.g. *Phalaris arundinaceae* (Lavergne & Molofsky 2007)] genetic diversity in the introduced range if a species had been introduced on several occasions or from several sources, as increased propagule pressure leads to high genetic diversity (Gentoni *et al*. 2005). However we do admit that low genetic diversity is not exclusively associated with natural dispersal, as a single human-mediated introduction or a single source population might also explain the low genetic diversity observed.

Our results suggest that *C. pechuelii* is an evolutionary consequence of novel environmental conditions acting on tropical/sub-tropical *C. corindum*. Southern African *C. corindum* and *C. pechuelii* are more closely related to each other than to *C. corindum* from South America based on the haplotype network and phylogeny (discussed above) (Fig. 3). Also they share more genetic material with each other than with *C. corindum* from South America (Fig. 2), are genetically clustered together (Fig. 4) and have least genetic structure between them (Table 1D). These results very clearly illustrate that *C. corindum* from southern Africa and *C. pechuelii* are closely related and that one evolved from the other.

The mixed monophyletic clade including southern African *C. corindum* and *C. pechuelii* is well supported (PP=1) with two clades further weakly supported, one including only *C. pechuelii* samples (P=0.72) and the other including both species with all but one *C. pechuelii* sample basal to *C. corindum* samples (PP=0.74) (Fig. 3). Posterior probabilities between 0.9 and 1 are well supported, between 0.7 and 0.89 moderately supported and lower 0.7 weakly supported (Schuster *et al*. 2013). Thus, while structure within the *C. corindum* and *C. pechuelii* clade is only moderately supported, it is still worthwhile to discuss. There are three scenarios to explain the mixed *C. corindum* and *C. pechuelii* clade. First, speciation of *C. pechuelii* occurred very recently and not enough evolutionary time has passed for complete genetic differentiation to occur between these two southern African species. Secondly, populations of *C. pechuelii* and *C. corindum* had secondary contact and are capable to intrebreed freely (hybridize), in which case gene flow is/was still present recently, preventing complete differentiation. The third possibility is that southern African *C. corindum* evolved from *C. pechuelii* following dispersal of *C. pechuelii* deeper into Namibia/southern Africa. In this case *C. pechuelii* would be more closely related to South American accessions than southern African *C. corindum* would be to South

American *C. corindum*, which is the case when considering the pairwise population structure (Table1 C).

The mixed monophyletic clade for *C. corindum* from southern African and *C. pechuelii* makes this species complex a promising system to investigate various species concepts. Species concepts have been debated for decades and are still disputed with approximately 26 different species concept definitions (Wilkins 2011). The lack of agreement between scientists on species concepts often poses problems for identifying species. If southern African *C. corindum* indeed evolved from *C. pechuelii* and not the other way around, then this taxonomy should be reviewed, at least based on the genetic and possibly biological species concept. A quick look at *C. corindum* from South America and southern Africa won't reveal any immediate morphological difference, however more indepth analsys on the morphology of plants in southern Africa and South America might reveal minor differences to support the genetic difference observed between the two continents.

The evolutionary diversification of *C. pechuelii* from *C. corindum* is also a remarkable example of how rapidly diversification can occur. In chapter 2 the arrival or *C. corindum* to southern Africa have been estimated no later than ca. 5.5 MYA. Therefore the evolution of a subtropical *C. corindum* to a desert *C. pechuelii* must have occurred within the last 5.5 million years, in all likelihood more recently. This rapid diversification is not unexpected with a previous study finding *Cardiospermum* and *Paullinia* species to have significantly higher evolutionary rates compared to other Sapindaceae genera (Harrington 2008).

3.6 Conclusion

We show the peculiar evolution of a desert species following long distance dispersal from South America, illustrating how novel environmental conditions can lead to evolution. We

couldn't however determine the direction of evolution between *C. corindum* from southern Africa and *C. pechuelii*. To further unravel relationships between these species and the path of evolution, morphological and additional genetic analyses will have to be incorporated, making this system promising to investigate various species concept.

4. Thesis conclusions

The large-scale movements of balloon vine species have in some instances distorted the group's historical biogeography. For example, at least one species (*C. grandiflorum*) has been described as a known invader in southern Africa, Australia and other parts of the world, while the statuses of *C. halicacabum* and *C. corindum* have largely been debated. In South Africa the extent of *C. grandiflorum* invasions lead to the government's 'Working for Water' program launching a biological control research initiative against this species. However, none of the numerous control agents identified over the last decade have been released due to concerns about potential non-target impacts on *C. corindum* and *C. halicacabum* given their unknown native statuses.

In Chapter one I reviewed and highlighted the historical movements of *Cardiospermum* species and how this contributed to numerous biogeographic uncertainties. Species distribution modelling results showed that *Cardiospermum* species have the potential to spread further in regions in Australia, Africa and Asia, underlining the importance of resolving taxonomic uncertainties to improve management efforts. I also concluded that, species distribution modelling over-fit native range data prediction of suitable ranges, and thus factors other than climate influence establishment potential.

In Chapter two, using a phylogenetic approach, I resolved some of the biogeographic issues identified for some *Cardiospermum* species. Specifically, I illustrated that *C. halicacabum* is non-native in southern Africa while *C. corindum* is native due to natural long distance dispersal. I also recommended that the alien status of *C. grandiflorum* be reviewed in Hawaii and Uganda due to a deep node split from South American samples in the phylogeny and haplotypes several mutational steps away from South American haplotypes. An unexpected finding was the close phylogenetic

relationship observed between southern African *C. pechuelii* and *C. corindum* which prompted further investigation as outline below.

The observed close relationship between *C. corindum* and *C. pechuelii* was investigated in chapter three. I concluded that *C. pechuelii* rapidly evolved from *C. corindum* likely following a single long distance dispersal event. The haplotype network and phylogeny showed that *C. pechuelii* and *C. corindum* from southern Africa are more closely related to each other than to *C. corindum* from South America. However the mixed monophyletic clade for southern African *C. corindum* and *C. pechuelii* blurred evolutionary history, which makes this a promising system to test different species concepts.

Overall, my thesis gives insights into the biology and ecology of the cosmopolitan genus *Cardiospermum*. Caution should be exercised when introducing these species as ornamental or medicinal plants since they show significant potential for further distribution. Vigilance against *C. corindum* introduction in Australia is especially important since the majority of Australia seems highly suitable. The *C. corindum*/*C. pechuelii* species complex would be an interesting system to test different species concepts as unresolved evolutionary events characterise the relationship between southern African species. To conclude on a biological control agent against invasive *C. grandiflorum* after 10 years of research in South Africa, I strongly advise that current identified biological control agents not be released due to potential non-target impact on native species *C. corindum* and *C. pechuelii*. Future research regarding biological control should identify host specific agents for *C. grandiflorum* and possibly *C. halicacabum*. We also advise further research to review the invasive/alien status awarded to *C. grandiflorum* in tropical Africa and Hawaii as our results indicated a possible native status. My study clearly illustrates the importance of biogeographical information when identifying native ranges and developing biological control research programmes.

References

Adeyemi TO, Ogundipe OT (2012) Biodiversity of Sapindaceae in West Africa: a checklist. *International Journal of Biodiversity and Conservation*. 4: 326-331.

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and true skill statistic (TSS). *Journal of Applied Ecology*. 43: 1223-1232.

Araújo M B, Luoto M (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*. 16: 743-753.

Arthofer W (2010) tinyFLP and tinyCAT: software for automatic peak selection and scoring of AFLP data tables. *Molecular Ecology Resources*. 10: 385-388.

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997). Evolution of coalescence times, genetic diversity and structure during colonisation. *Theoretical Population Biology.* 51: 148–164.

Baldwin BG, Sanderson MJ (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Sciences*. 95: 9402–9406.

Bartish IV, Antonelli A, Richardson JE, Swenson U (2010) Vicariance or long-distance historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography.* 38: 177-190.

Bean AR (2007) A new system for determining which plant species are indigenous in Australia. *Australian Systematic Botany*. 20: 1-43.

Blignaut M, Ellis AG, Le Roux JJ (2013) Towards a Transferable and Cost-Effective Plant AFLP Protocol. *PLOS ONE*. 8: e61704.

Bowen B, Johnson K, Franklin S, Call G, Webber M (2002) Invasive exotic pest plants in Tennessee. *Journal of the Tennessee Academy of Science*. 77: 45-48.

Brighenti AM, Voll E, Gazziero DLP (2003) Biology and management of *Cardiospermum halicacabum*. *Planta Daninda*. 21: 229-237.

Brizicky GK (1963) The genera of Sapindales in the Southeastern United States. *Journal of the Arnold Arboretum*. 44: 462-501.

Broennimann O, Treier UA, Müller-Schärer H, Thiuller W, Peterson AT, Guisan A (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*. 10: 701–709.

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*. 21:255-265.

Burke A (2003) Floristic relationships between inselbergs and mountain habitats in the central Namib. *Dinteria*. 28: 19-38.

Carlquist S, Baldwin BG, Carr GD (2003) Tarweeds and silverswords: evolution of the Madiinae (Asteraceae). Missouri Botanical Garden Press, Saint Louis, MO.

Carroll SP (2007) Natives adapting to invasive species: ecology, genes, and the sustainability of conservation. *Ecological Research*. 22: 892-901.

Carroll SP (2011) Conciliation biology: the eco-evolutionary management of permanently invaded biotic systems. *Evolutionary Applications*. 4: 184-199.

Carroll SP, Boyd C (1992) Host race radiation in the soapberry bug: Natural history with the history. *Evolution*. 46: 1052-1069.

Carroll SP, Mathieson M, Loye J (2005a) Invasion history and ecology of the environmental weed balloon vine, *Cardiospermum grandiflorum* Swartz, in Australia. *Plant Protection Quarterly*. 20: 140-144.

Carroll SP, Loye JE, Dingle H, Mathieson M, Famula TR, Zalucki P (2005b) And the bleak shall inherit - evolution in response to invasion. *Ecology Letters*. 8: 944-951.

Carroll SP, Loye JE (2012) Soapberry bug (Hemiptera: Rhopalidae: Serinethinae) native and introduced host plants: the biogeographic background of anthropogenic evolution. *Annals of the Entomological Society of America*. 105: 671-684.

Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology, Evolution, and Systematics*. 15: 97–131.

Castellanos AE, Tinoco-Ojanguren C, Molina-Freaner F (1999) Microenvironmental heterogeneity and space utilization by desert vines within their host trees. *Annals of Botany*. 84: 145–153.

Cavender-Bares J, Kozak KH, Fine PVA, Kembel ST (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*. 12: 693-715.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*. 9: 1657–1660.

Cowie ID, Finlayson CM (1986) Plants of the Alligator Rivers Region Northern Territories. Government Publishing Service Canberra, Canberra, Australia.

Davies FG, Verdcourt B (1998) Flora of tropical East Africa: Sapindaceaea. HJ Beentjie and CM Whitehouse (Eds) CRC Press, Kew: 97-102.

Davis MA, Grime JP, Thompson K (2000) Fluctuating resources in plant communities: a general theory of invasibility. *Journal of Ecology*. 88: 528–534.

Dempsey MA (2011) Anatomical and morphological responses of *Cardiospermum halicacabum* l. (balloon vine), to four levels of water availability. MSc Dissertation University of North Texas.

Doyle JJ, Doyle JL (1990) A rapid total DNA preparation procedure for fresh plant tissue. *Focus*. 12: 13-15.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) A Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*. 29: 1969-1973.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 4: 359-361.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*. 77: 802-813.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*. 7: 574-578.

Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Letters*. 27: 861-874.

Ferrucci MS, Acevedo-Rodrigues P (1998) *Cardiospermum cuchujaquense* (Sapindaceae), a new species from Sonora, Mexico. *Novon*. 8: 235-238.

Ferrucci MS, Umdiriri JD (2011) *Cardiospermum bahianum* (Sapindaceae: Paullinieae), a new species from Bahia, Brazil. *Systematic Botany*. 34: 950-956.

Ferrucci MS, Urdampilleta JD (2011) *Cardiospermum cristobaliae* (sapindaceae, Paullinieae), una nueva especie de Minas Gerais, Brasil. *Brittonia*. 63: 478-483.

Flora of China. Online Available: http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=200013187 Last accessed 23 Jan 2013.

Flora of New Zealand. 2012. Online Available: http://www.nzflora.info/factsheet/Weed/Cardiospermum_grandiflorum.html Accessed 15 Jan 2013.

FloraBase. Online Available: http://florabase.dec.wa.gov.au/browse/profile/17318 Last accessed 15 Jan 2012.

Fujita S, Sends Y, Nakaguchi S, Hashimoto T (2001) Multiplex PCR using internal transcribed spacer 1 and 2 regions for rapid detection and identification of yeast strains. *Journal of Clinical Microbiology*. 39: 3617–3622.

GBIF - Global Biodiversity Information Facility (2010) Free and open access to biodiversity data. http://data.gbif.org/welcome.htm Last accessed 15 Jan 2012.

Gentoni BJ, Shykoff JA, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology*. 14: 4275-4285.

Germishuizen G, Meyer NL, Steenkamp Y, Keith M (2006) A checklist of South African plants. Southern African Botanical Diversity Network Report No. 41, SABONET, Pretoria.

Gildenhuys E, Ellis A, Carroll S, Le Roux JJ (2013) The ecology, biogeography, history and future if two globally important weeds: *Cardiospermum halicacabum* Linn. and *C. grandiflorum* SW. *Neobiota*.

Givnish TJ, Millam KC, Evans TM, Hall JC, Pires JC, Berry PE, Sytsma KJ (2004) Ancient vircariance or recent long-distance dispersal? Inferences about phylogeny and South American-African disjunctions in Rapateaceae and Bromeliaceae based on ndhF sequence data. *Internaional Journal of Plant Science.* 165: S35-S54.

Goolsby JA, De Barro PJ, Makinson JR, Pemberton RW, Hartley DM, Frohlich DR (2006) Matching the origin of an invasive weed for selecting a herbivore haplotype for a biological control programme. *Molecular Ecology*. 15: 287-297.

Goosem S (2008) Invasive weeds in the wet tropics. In: Stork NE, Turton SM (Eds) Living in a dynamic tropical forest landscape. Blackwell Publishing (Oxford): 311-316.

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*. 8: 993–1009.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41: 95-98.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 143: 29-36.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*. 2: 618-620.

Harrington M (2008) Phylogeny and evolutionary history of Sapindaceae and Dodonaea. PhD thesis. James Cook University.

Harrington MG, Gadek PA (2009). A species well travelled – the *Dodonaea viscosa* (Sapindaceae) complex based on phylogenetic analyses of nuclear ribosomal ITS and ETSf sequences. *Journal of Biogeography*. 36: 2313-2323.

Harris CJ, Murray BR, Hose GC, Hamilton MA (2007) Introduction history and invasion success in exotic vines introduced to Australia. *Diversity and Distribution*. 13: 467-475.

Henderson L (2001) Alien weeds and invasive plants. Agricultural Research Council (Cape Town): 60-61.

Henderson L (2007) Invasive, naturalized and casual alien plants in southern Africa: a summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*. 37: 215–248.

Henry RD, Scott AR (1981) Time of introduction of the alien component of the spontaneous Illinois vascular flora. *American Midland Naturalist*. 106: 318-324.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*. 25: 1965-1978.

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*. 199: 142-152.

Hoffmann AA, Parsons PA (1997) Extreme Environmental Change and Evolution. Cambridge University Press, Cambridge.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. 23: 254-267.

Hyde M, Wursten B, Ballings P (2012a) Flora of Zimbabwe. http://www.zimbabweflora.co.zw/ Last accessed 15 Jan 2012.

Hyde M, Wursten B, Ballings P, Dondeyne S (2012b) Flora of Mozambique. http://www.mozambiqueflora.com/ Last accessed 15 Jan 2012.

James EP (1825) Catalogue of plants collected during a journey to and from the rocky mountains, during the summer of 1820. *Transactions of the American Philosophical Society*. 2: 172-190.

Jiménez-Valverde A, Lobo JM, Hortal J (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distribution*. 14: 885-890.

Johnston KS, Murray DS, Williams JC (1979) Germination and emergence of balloon vine (*Cardiospermum halicacabum*). *Weed Science*. 27: 73-76.

Jolley ER, Walker RH, McGuire JA, Johnston SK, Murray DS, Williams JC (1983) Balloonvine biology and control in soybeans. *Alabama agricultural experiment station, Auburn University*, 547:1-36. http://repo.lib.auburn.edu/repo/bitstream/handle/123456789/2468/1726BULL.pdf?sequence=1

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*. 30: 772-780.

Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution*. 17: 164-170.

Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species ranges. *Ecology Letter*. 12: 1-17.

Knope ML, Morden CW, Funk VA, Fukami T (2012) Area and the rapid radiation of Hawaiian Bidens (Asteraceae). *Journal of biogeography*. 39: 1206-1217.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*. 104: 3883-3888.

Le Roux JJ, Wieczorek AM, Ramadan MM, Tran CT (2006) Resolving the native provenance of invasive fireweed (*Senecio madagascariensis Pior*.) in the Hawaiian Islands as inferred from phylogenetic analysis. *Diversity and Distribution*. 12: 694-702.

Le Roux JJ, Wieczorek AM (2009) Molecular systematics and population genetics of biological invasions: towards a better understanding of invasive species management. *Annals of Applied Biology*. 154: 1- 17.

Lexer C, Fay MF (2005) Adaptation to environmental stress: a rare or frequent driver of speciation? *Journal of Evolutionary Biology*. 18: 893-900.

Ley AC, Hardy OJ (2013) Improving AFLP analysis of large scale patterns of genetic variation – a case study with the Central African lianas *Haumania* spp (Marantaceae) showing interspecific gene flow. *Molecular Ecology*. 22: 1984-1997.

Llewellyn CF, Richardson DM, Wilson SRU (2008) Ornamental Plants as Invasive Aliens: problems and solutions in Kruger National Park, South Africa. *Environmental Management*. 41: 32–51.

Mc Kay F, Oleiro M, Fourie A, Simelane D (2010) Natural enemies of balloon vine *Cardiospermum grandiflorum* (sapindaceae) in Argentine and their potential use as biological control agents in South Africa. *International Journal of Tropical Insect Science*. 30: 67-76.

Messing RH, Wright MG (2006) Biological control of invasive species: solution or pollution? *Frontiers in Ecology and the Environment*. 4: 132–140.

Meyer JY (2004) Threat of invasive alien plants to native flora and forest vegetation of Eastern Polynesia. *Pacific Science*. 58: 357–375.

Molina-Freaner F, Tinoco-Ojanguren C (1997) Vines of a desert plant community in central Sonora, Mexico*. Biotropica*. 29: 46–56.

Mosango M, Maganyi O, Namaganda M (2001) A floristic study of weed species of Kampala (Uganda). *Systematics and geography of plants*. 71: 223-236.

Murty PP, Venkaiah M (2011) Biodiversity of weed species in crop fields of north coastal Andhra Pradesh, India. *Indian Journal of Fundamental and Applied Life Sciences*. 1: 59-67.

Ndlovu J, Richardson DM, Wilson JRU, O'Leary M, Le Roux JJ (2013) Elucidating the native sources of an invasive tree species, *Acacia pycnantha*, reveals unexpected native range diversity and structure. *Annals of Botany*. 5: 895-904.

Nettel A, Dodd RS (2007) Drifting propagules and receding swamps: Genetic footprints of mangrove recolonization and dispersal along tropical coasts. *Evolution*. 61: 958-971.

Nogales M, Heleno R, Traveset A, Vargas P (2012) Evidence overlooked mechanisms of long-distance seed dispersal to and between oceanic islands. *New Phytologists*. 194: 313-317.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6: 288- 295.

Perreira C, Carroll S, Loye J (2012) *Leptocoris ursulae*, a new species of soapberry bug from Uganda (Heteroptera: Rhopalidae, Serinethinae). *Entomologische Zeitschrift*. 122: 123-124.

Petitpierre B, Kueffer C, Broennimann O, Randin C, Daehler C, Guisan A (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science*. 335: 1344-1347.

PIER, Pacific Island Ecosystems at Risk (Last updated 2013). Online Available: http://www.hear.org/pier/index.html Last accessed 15 Aug 2013.

Pimentel D, Lach L, Zuniga R, Morrison D (2000) Environmental and economic costs associated with non-indigenous species in the United States. *BioScience*. 50: 53–65.

Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, Wong E, Russel L, Zern J, Aquino T, Tsomondo T (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems & Environment*. 84: 1–20.

Prentis PJ, Sigg DP, Raghu S, Dhileepan K, Pavasovic A, Lowe AJ (2009) Understanding invasion history: genetic structure and diversity of two globally invasive plants and implications for their management. *Diversity and Distributions*. 25: 822-830.

Price JP, Wagner WL (2004) Speciation in Hawaiian angiosperm lineages: cause, consequence, and mode. *Evolution*. 58: 2185–2200.

Pulliam HR (2000) On the relationship between niche and distribution. *Ecology letters*. 3: 349-361.

Pysek P, Richardson DM (2010) Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*. 35: 25-55.

Quinn LD, Matlaga DP, Stewart JR, Davis AS (2011) Emperical evidence of long-distance dispersal in *Micanthus sinensis* and *Miscanthus x giganteus. Invasive plant science and management*. 4: 142-150.

R Development Core Team (2012) A language and environment for statistical computing. R Foundation for Statistical Computing.

Raju AJS, Ramana KV, Roa NG, Varalakshmi P (2011) Monoecy and entomophily in *Cardiospermum canescens* Wall. (Sapindaceae), a medicinally valuable herbaceous vine. *Current Science*. 101: 617-619.

Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Redfern JL (2012) Phylogeny and phylogeography of Fouquieriaceae: Evolutionary history of a North American warm desert plant family. Online Available: http://udini.proquest.com/view/phylogeny-and-phylogeography-of-goid:304525559/ Last accessed 15 Jan 2013.

Renner S (2004) Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Science*. 165: S23-S33.

Rock BN (1972) The Woods and Flora of the Florida Keys: "Pinnatae". *Smithsonian Contributions to Botany*. 5: 1-35.

Roderick GK, Navajas M (2003) Genes in new environments: genetics and evolution in biological control. *Nature Reviews Genetics*. 4: 889-899.

Ronquist F, Huelsenbeck JP (2003) Mr Bayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19: 1572–1574.

Sakai AK, Allendorf FW, Holt JS, Lodge MD, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*. 32: 305-332.

Saltonstall K (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences*. 99: 2445-2449.

SANBI (South African National Biodiversity Institute) boidiversity of live. http://redlist.sanbi.org/species.php?species=3845-2 Last accessed 04 Dec 2012.

Save Our Waterways Now. Online available: http://www.saveourwaterwaysnow.com.au/01_cms/details_pop.asp?ID=342 Last accessed 30 Nov 2013

Schaal BA, Gaskin JF, Caicedo AL (2003) Phylogeography, haplotype trees, and invasive plant species. *Journal of Heredity*. 94: 197-204.

Schuster TM, Setaro SD, Kron KA (2013) Age Estimates for the Buckwheat Family Polygonaceae Based on Sequence Data Calibrated by Fossils and with a Focus on

the Amphi-Pacific *Muehlenbeckia*. PLOS one. 8: e61261. doi:10.1371/journal.pone.0061261.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare. *American Journal of Botany*. 94: 275-288.

Simelane DO, Fourie A, Mawela KV (2011) Prospective agents for the biological control of *Cardiospermum grandiflorum* Sw (sapindaceae) in South Africa. *African Entomology*. 19: 269-277.

Sinclair SJ, White MD, Newell GR (2010) How useful are species distribution models for managing biodiversity under future climates? *Ecology and Society*. 15: 8.

Soons MB, Bullock JM (2008) Non-random seed abscission, long-distance wind dispersal and plant migration rates. *Journal of Ecology*. 96: 581-590.

Subramanyam R, Newmaster SG, Paliyath G, Newmaster CB (2007) Exploring ethnobiological classifications for novel alternative medicine: A case study of *Cardiospermum halicacabum L.* (Modakathon, Balloon Vine) as a traditional herb for treating rheumatoid arthritis. *Ethnobotany*. 19: 1-18.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*. 17: 1105–1109.

Takayama K, Tamura M, Tateishi Y, Webb EL, Kajita T (2013) Strong genetic structure over the American continents and transoceanic dispersal in the mangrove genus Rhizophora (Rhizophoraceae) revelaed by broad-scale nuclear and chloroplast DNA anlysis. *American Journal of Botany*. 100: 1191-1201.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. 28: 2731-2739.

Templeton AR (2008) The reality and importance of founder speciation in evolution. B*ioEssays*. 30: 470-479.

The Plant List (2010) Version 1. Available Online: http://www.theplantlist.org/ last accesses 05 Dec 2012

Thompson K, Hodgson JG, Grime JP, Burke MJW (2004) Plant traits and temporal scale: evidence from a 5-year invasion experiment using native species. *Journal of Ecology*. 89: 1054-1060.

Thuiller W, Richardson DM, Pysek P, Midgley GF, Hughes GO, Rouget M (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*. 11: 2234-2250.

Thuiller W, Lafourcade B, Engler R, Araujo MB (2009) BIOMOD - a platform for ensamble forecasting of species distributions. *Ecography*. 32: 369-373.

Tremetsberger K, Weiss-Shneeweiss H, Stuessy T, Samuel R, Kadlec G, Orits MA, Talavera S (2005) Nuclear ribosomal DNA and karyotypes indicate a NW African origin of South American *Hypochaeris* (Asteraceae, Cichorieae). *Molecular Phylogenetics and Evolution*. 35: 102–116.

Urdampilleta LD, Coulleri JP, Ferrucci MS, Forni-Martins ER (2012) Karyotype evolution and phylogenetic analyses in the genus *Cardiospermum* L. (Paullinieae, Sapindaceae). Plant biology ISSN 1435-8603.

USDA - United States Department of Agriculture (2012) http://www.usda.gov/wps/portal/usda/usdahome last accessed 04 Dec 2012.

Van Damme P (1991) Plant ecology of the Namib Desert. *Afrika Focus*. 7: 355-400.

Van Wilgen BW, Richardson DM, Le Maitre DC, Marias C, Magadlela D (2001) The economical consequences of alien plant invasions: examples of impacts and approaches to sustainable management in South Africa. *Environment, Development and Sustainability*. 3: 145-168.

Venkatesh Babu KC, Krishnakumari S (2006) *Cardiospermum halicacabum* suppresses the production of TNF-alpha and nitric oxide by human peripheral blood mononuclear cells. *African Journal of Biomedical Research*. 9: 95-99.

Voll E, Brighenti AM, Gazziero DLP, Adegas FS (2004) Population dynamics of *Cardiospermum halicacabum* and competition with soybeans. *Pesquisa Agropecuária Brasileira*. 39: 27-33.

Wagner WL, Funk VA (eds) (1995) Hawaiian biogeography: evolution on a hot spot archipelago. Smithsonian Institution Press, Washington, DC.

Weckerle CS, Rutishauser R (2005) Gynoecium, fruit and seed structure of Paullinieae (Sapindaceae). *Botanical Journal of the Linnean Society*. 147: 159–189.

Weeds of Australia, Biosecurity Queensland Edition. Balloon vine *Cardiospermum grandiflorum*. http://keyserver.lucidcentral.org/weeds/data/03030800-0b07-490a-8d04-0605030c0f01/media/Html/Cardiospermum_grandiflorum.htm Last accessed 12 Dec 2012.

Werren GL (2002) A bioregional perspective of weed invasion of rainforests and associated ecosystems: Focus on the Wet Tropics of north Queensland. In: Grice

AC and Setter MJ (Eds). Weeds of rainforests and associated ecosystems, Cooperative Research Centre for Tropical Rainforest Ecology and Management, (Queensland): 9-18.

West JG (1980) A taxonomic revision of Dodonaea (Sapindaceae) in Australia. PhD thesis, University of Adelaide, Australia.

Wilkins JS (2011) Philosophically speaking, how many species concepts are there? *Zootaxa*. 1765: 58-60.

Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology and Evolution*. 24: 136-144.

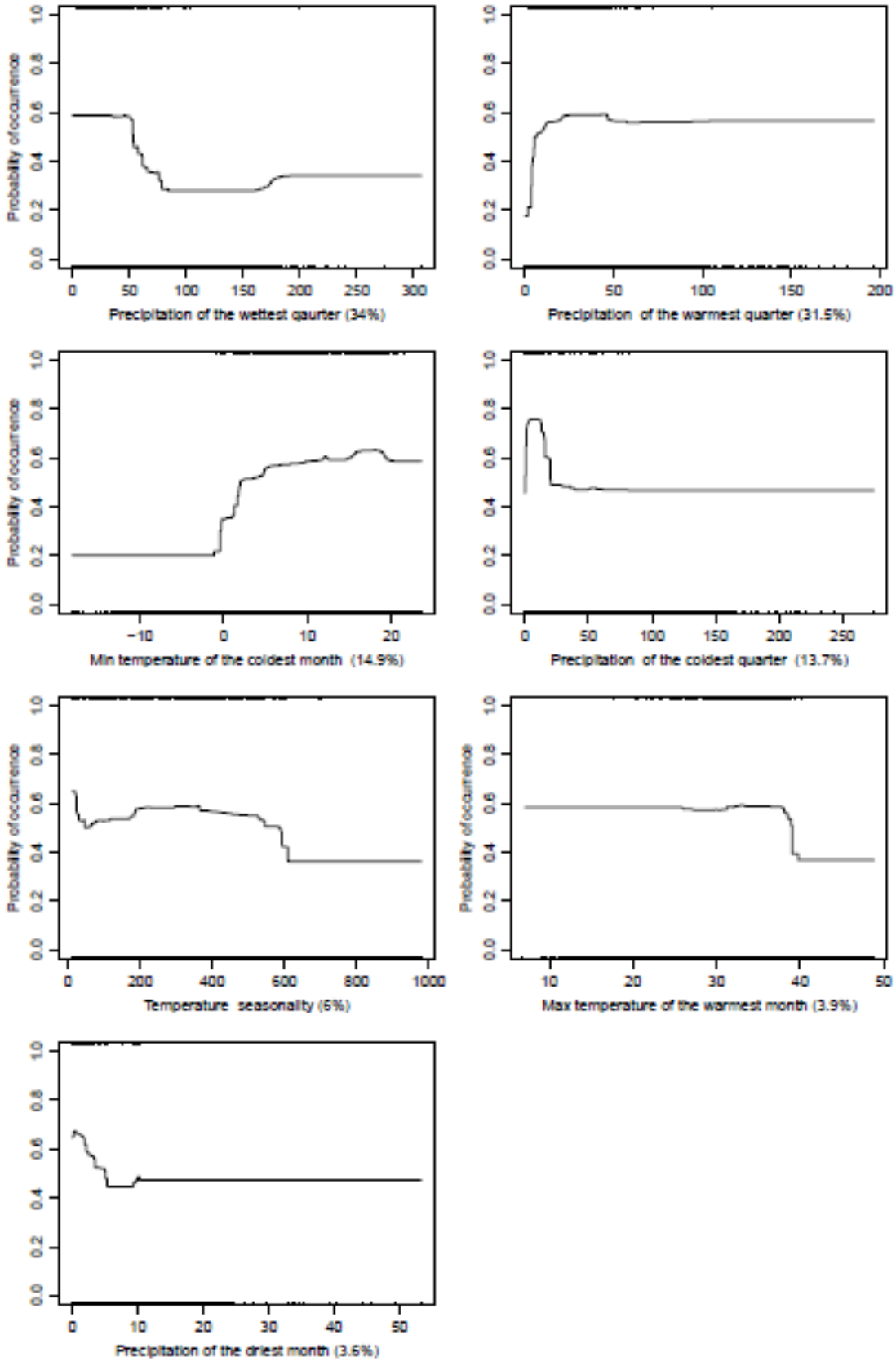
Supplementary material

Chapter 1

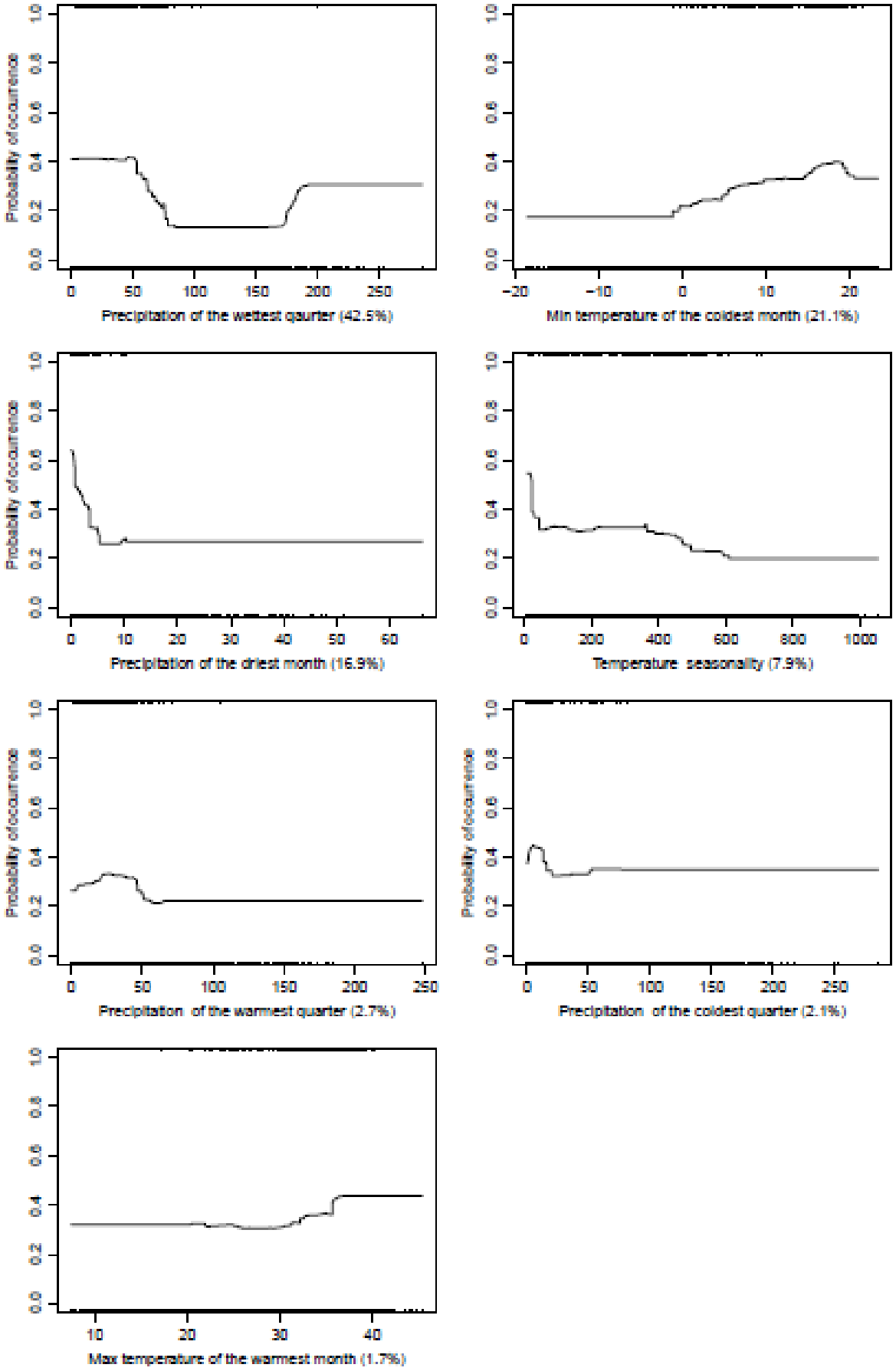
Supporting information for species distribution modelling of *Cardiospermum* species using native range presences and global pseudo absences:

Figure S1: Response plots of the variables used for modelling. Response plots for A) *C. corindum* global data, B) *C. corindum* native data, C) *C. grandiflorum* global data, D) *C. grandiflorum* native data, E) *C. halicacabum* global data and F) *C. halicacabum* native data. The plots are arranged according to the importance in brackets with the presence (top) and absence (bottom) points shown on the graph.

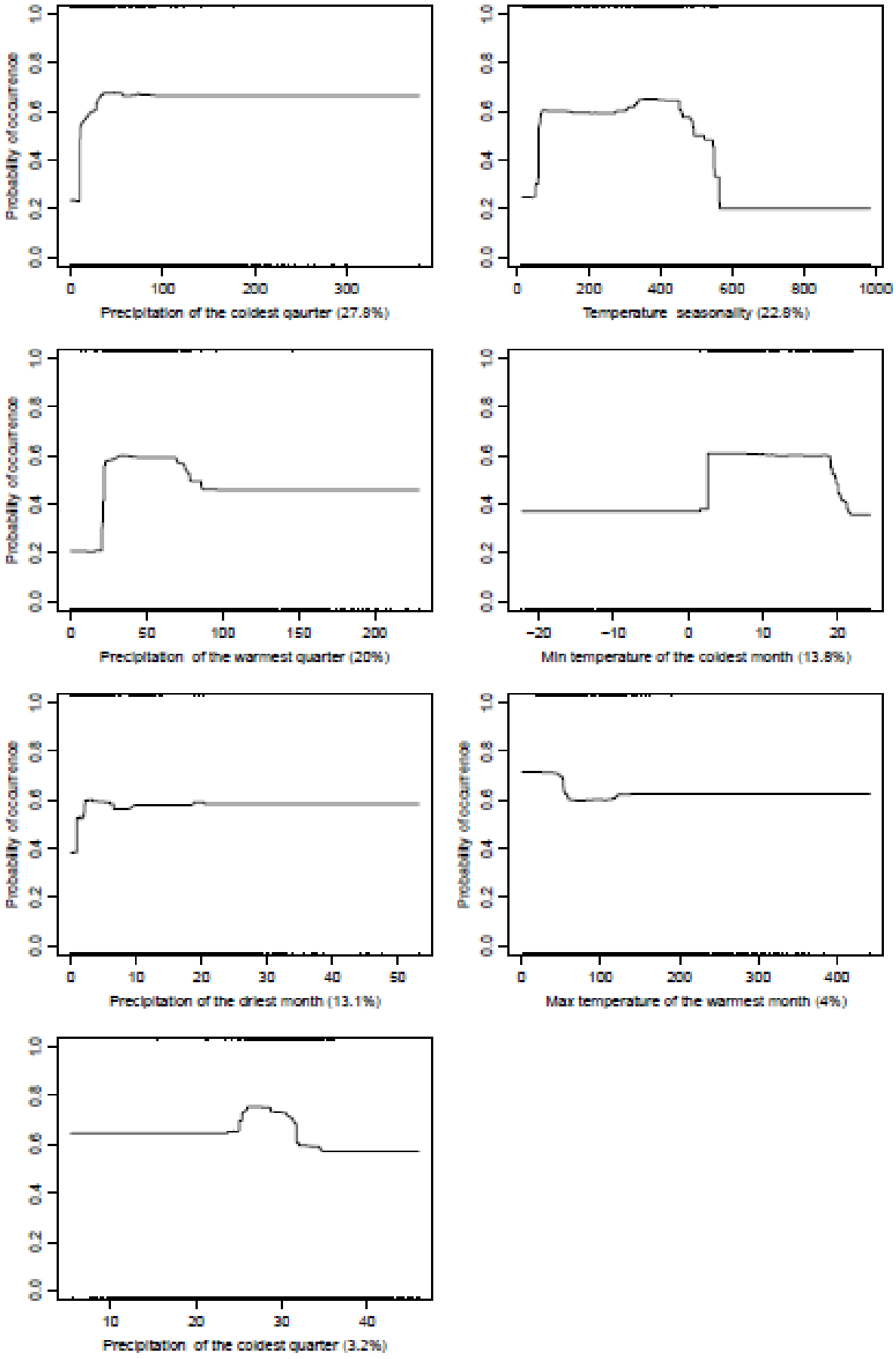
A


Probability of occurrence
Precipitation of the wettest qaurter (34%)
Precipitation of the warmest quarter (31.5%)
Min temperature of the coldest month (14.9%)
Precipitation of the coldest quarter (13.7%)
Temperature seasonality (6%)
Max temperature of the warmest month (3.9%)
Precipitation of the driest month (3.6%)

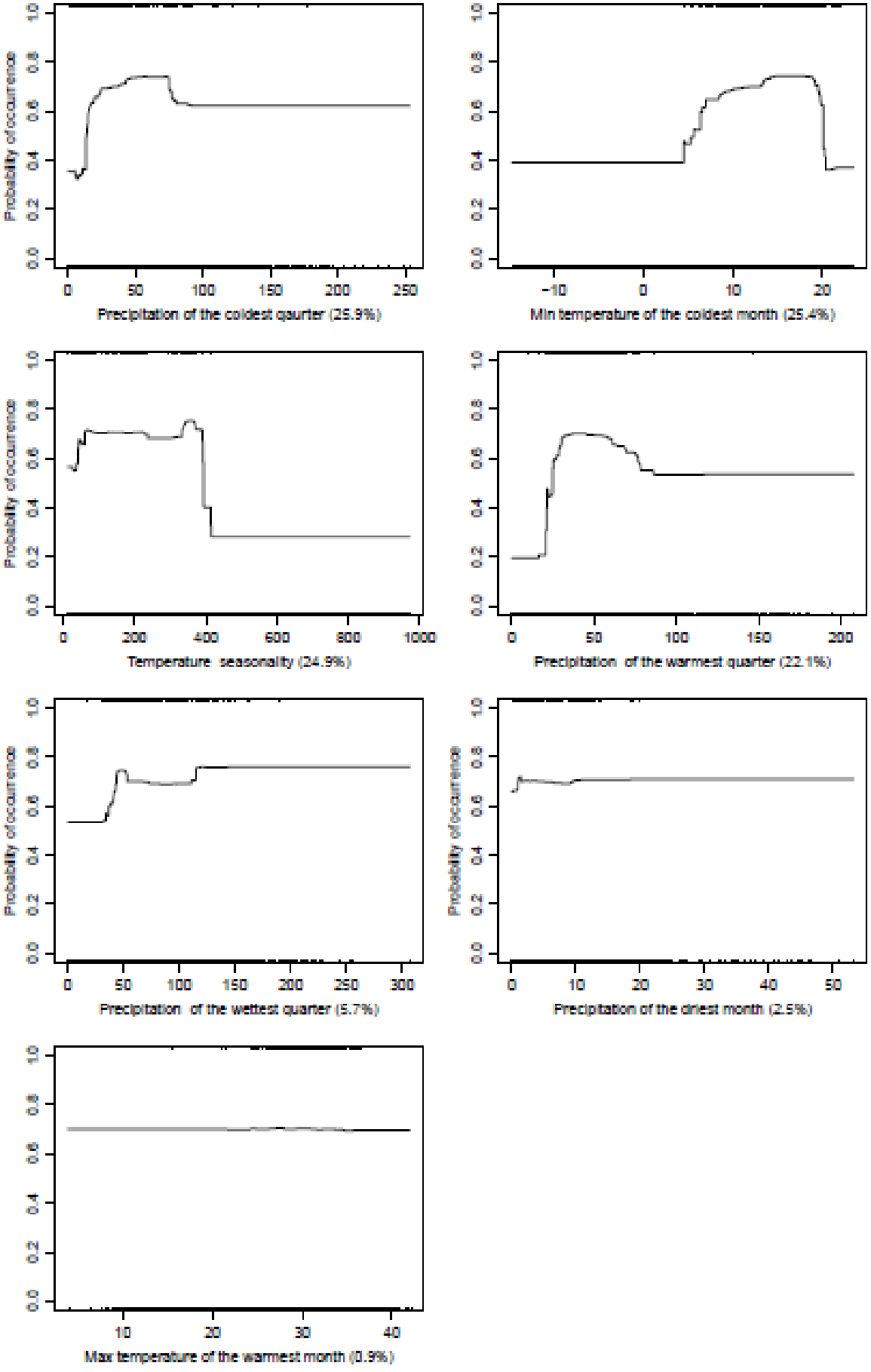
B

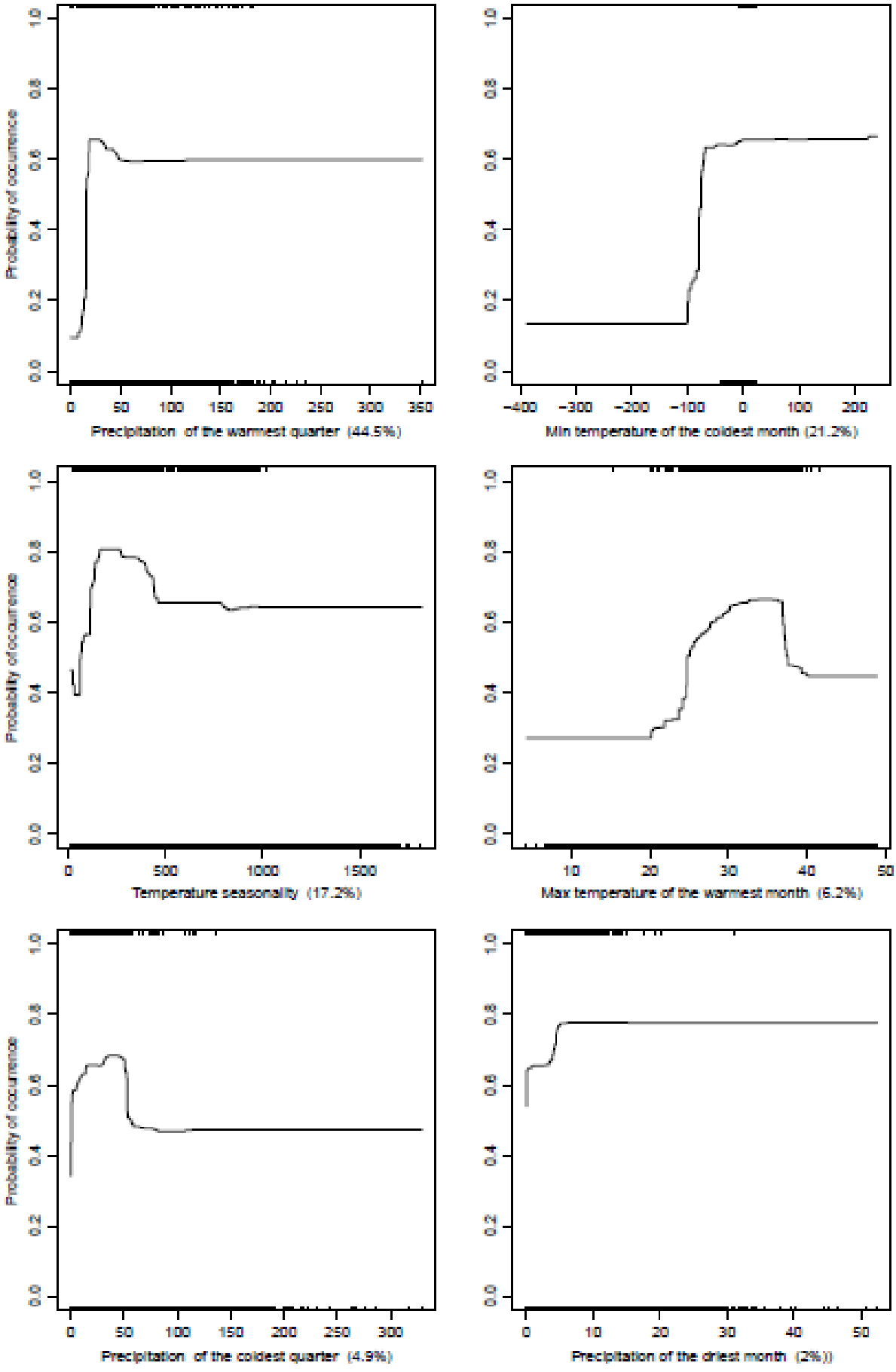

Probability of occurrence
Precipitation of the wettest qaurter (42.5%)
Min temperature of the coldest month (21.1%)
Precipitation of the driest month (16.9%)
Temperature seasonality (7.9%)
Precipitation of the warmest quarter (2.7%)
Precipitation of the coldest quarter (2.1%)
Max temperature of the warmest month (1.7%)

C

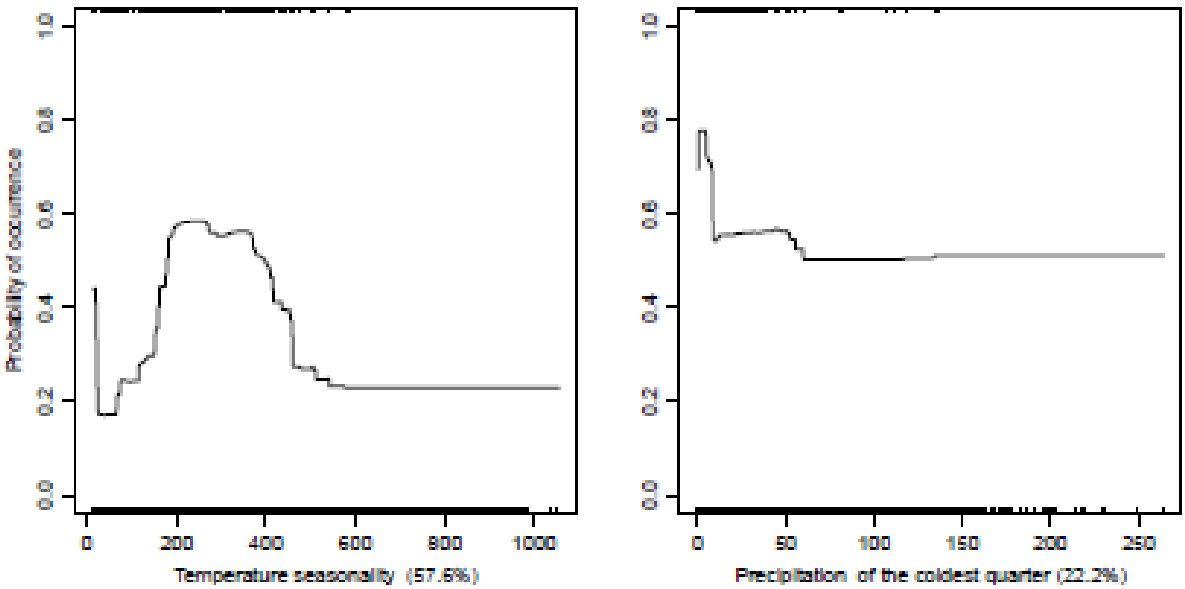

Probability of occurrence
Precipitation of the coldest qaurter (27.8%)
Temperature seasonality (22.8%)
Precipitation of the warmest quarter (20%)
Min temperature of the coldest month (13.8%)
Precipitation of the driest month (13.1%)
Max temperature of the warmest month (4%)
Precipitation of the coldest quarter (3.2%)

D

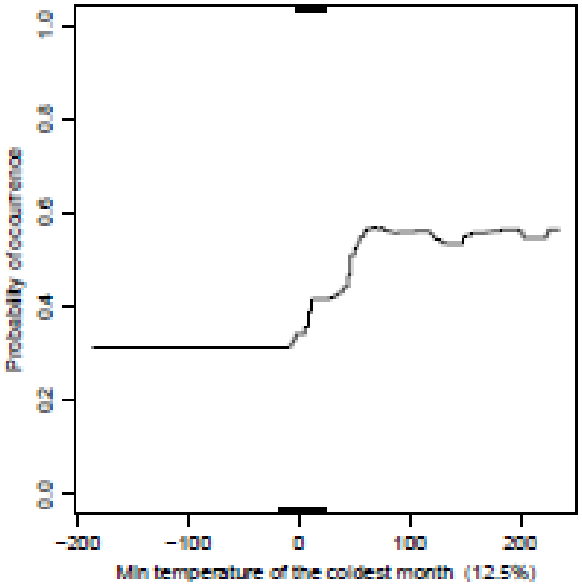

Probability of occurrence
Precipitation of the coldest qaurter (25.9%)
Min temperature of the coldest month (25.4%)
Temperature seasonality (24.9%)
Precipitation of the warmest quarter (22.1%)
Precipitation of the wettest quarter (5.7%)
Precipitation of the driest month (2.5%)
Max temperature of the warmest month (0.9%)

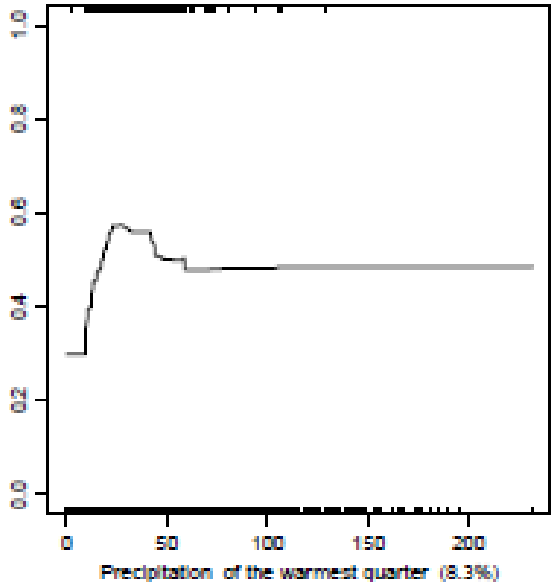

Probability of occurrence
Precipitation of the warmest quarter (44.5%)
Min temperature of the coldest month (21.2%)
Temperature seasonality (17.2%)
Max temperature of the warmest month (6.2%)
Precipitation of the coldest quarter (4.9%)
Precipitation of the driest month (2%))

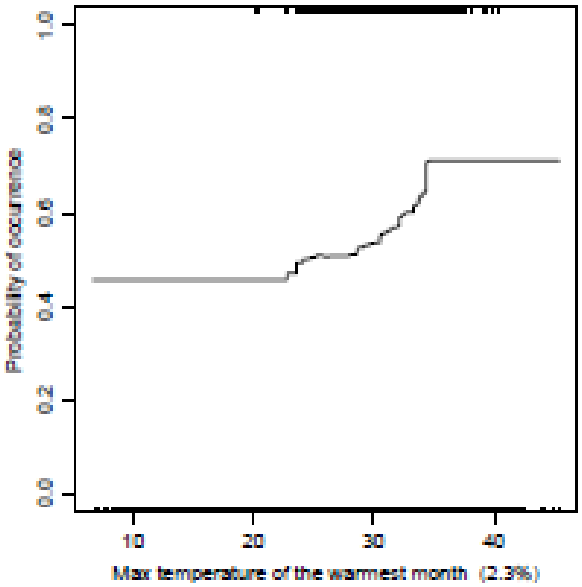
E

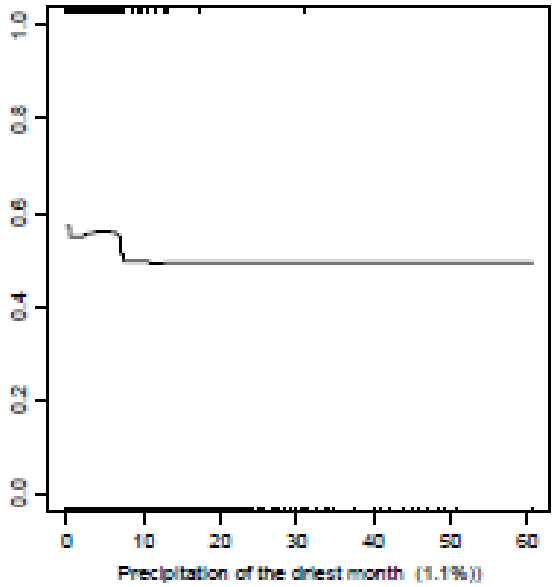

Probability of occurrence
Temperature seasonality (57.6%)

Precipitation of the coldest quarter (22.2%)


Probability of occurrence
Min temperature of the coldest month (12.5%)


Precipitation of the warmest quarter (8.3%)


Probability of occurrence
Max temperature of the warmest month (2.3%)


Precipitation of the driest month (1.1%))

F

Figure S2: Species distribution modelling predictions. Modelling predictions for A) *C. corindum* B) *C. grandiflorum* and C) *C. halicacabum* using native range presence data and pseudo absence points generated with global range data.

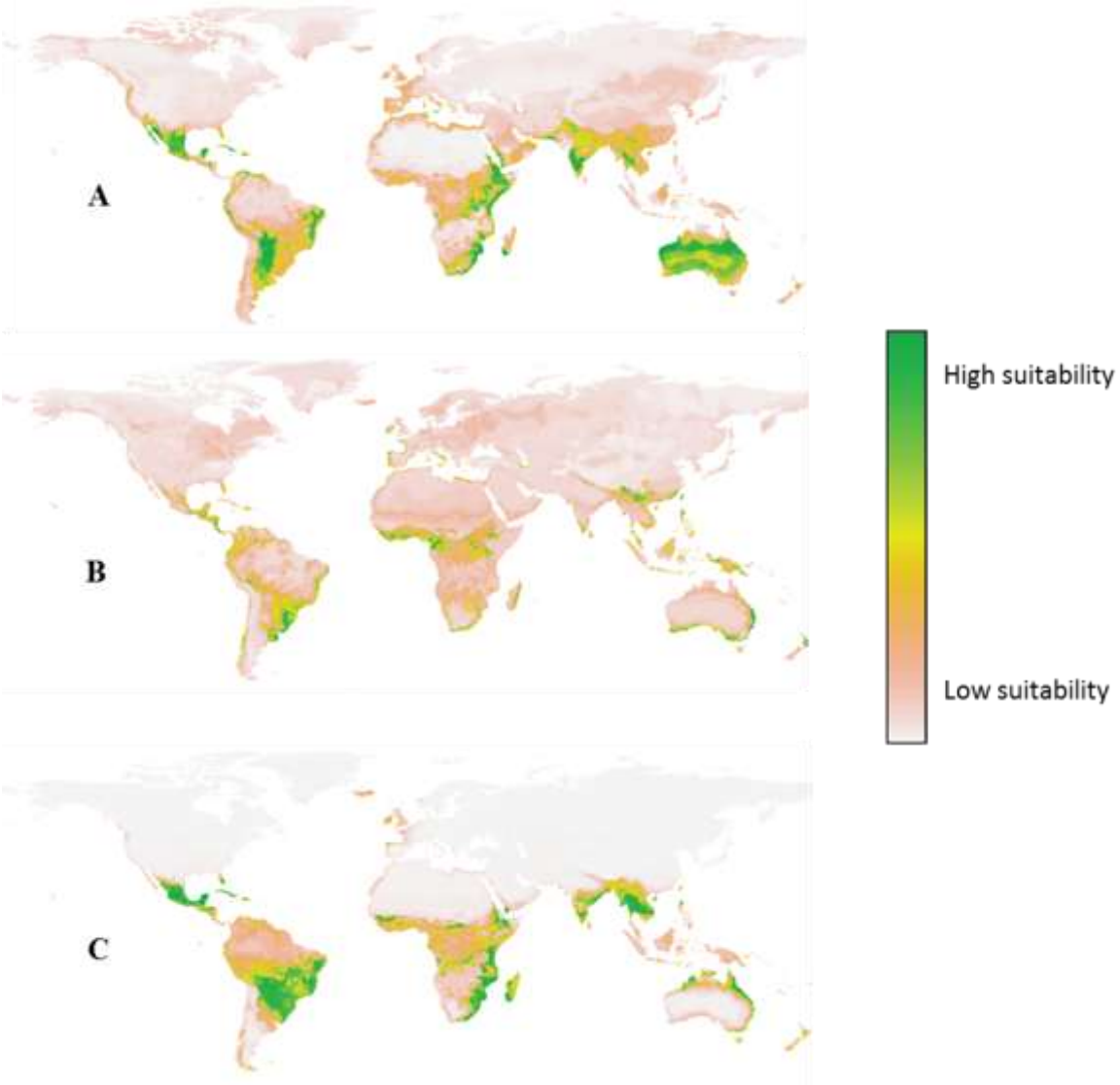


Table S1. Contribution of each BioClim variable for distribution modelling of three *Cardiospermum* species.

	Variable importance (%)		
Variable used for modelling	*C. halicacabum*	*C. grandiflorum*	*C. corindum*
Min temperature of the coldest month	0.9	17.2	13.3
Max temperature of the warmest month	2	30	0.9
Precipitation of the coldest quarter	10.2	24	63.8
Precipitation of the driest month	4.7	4.5	1.9
Precipitation of the warmest quarter	18.8	7.4	12.3
Temperature seasonality	74.2	2.2	9.6
Precipitation of the wettest quarter	-	35.9	22.5

Table S2. Evaluation of modelling predictions. True skill statistic (TSS) and area under the ROC curve (AUC) for native range presence data and global pseudo absence data, with independent data evaluation for the native range models.

Species	TSS	Independent data (TSS)	AUC	Independent data (AUC)
C. corindum	0.69	0.65	0.91	0.88
C. grandiflorum	0.72	0.57	0.93	0.85
C. halicacabum	0.77	0.46	0.95	0.79

Chapter 2

Figure S1: BEAST chronogram (based on *ITS* and *trnL-F*) with an exponential distribution at the root calibration, thus enforcing a younger tree for the genus *Cardiospermum*, including species with unknown native ranges *C. grandiflorum*, *C. corindum*, *C.*

halicacabum and the African species *C. pechuelii*. 95% confidence intervals for nodal ages are shown and scale bar indicating estimated time since divergence in millions of years.

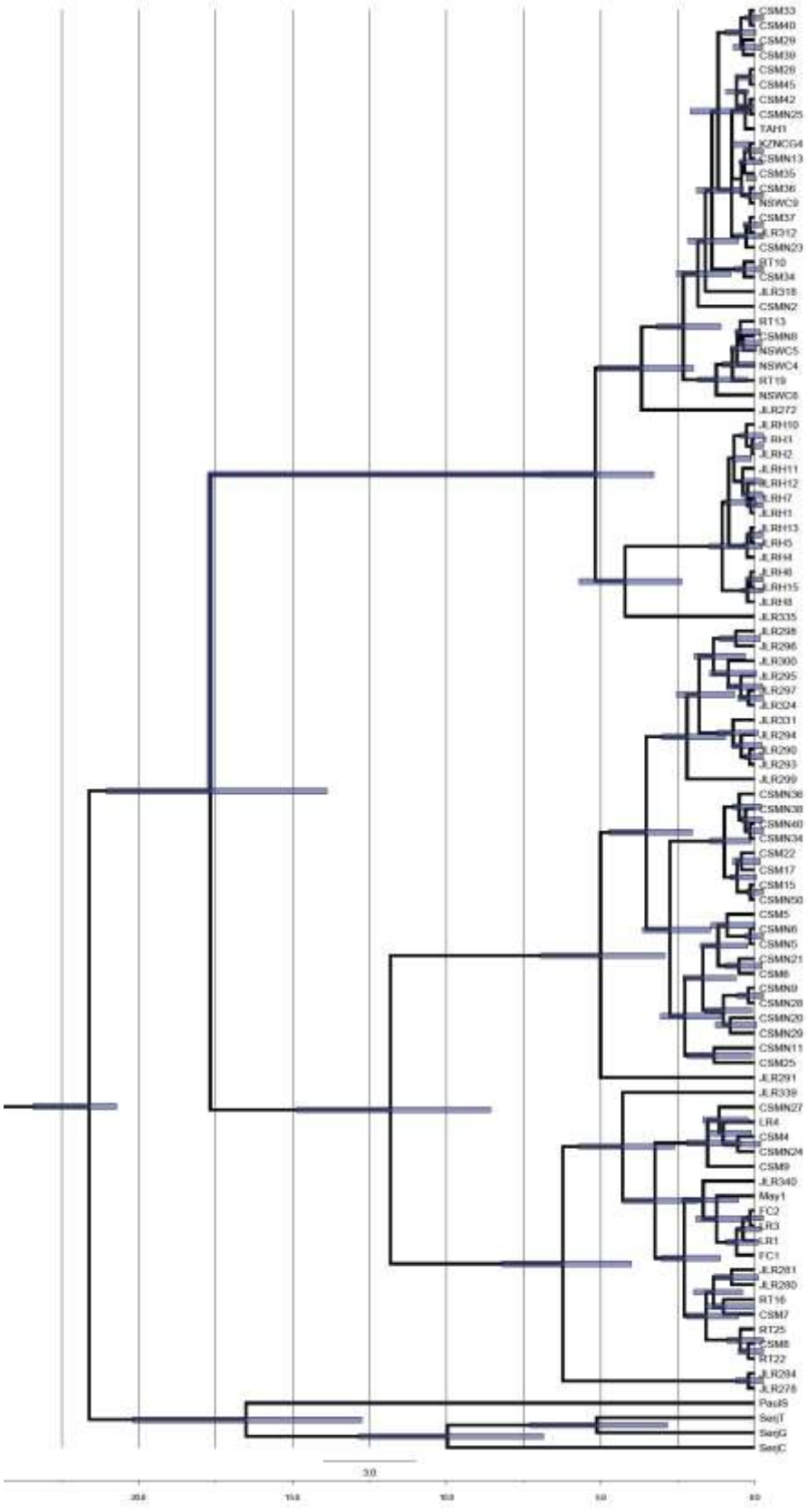


Table S1: Sample ID, species, country, latitude and longitude of samples used in this study.

Sample ID	Species	Country	latitude	longitude
CSM15	*C. pechuelii*	Namibia	-21.3760	15.0892
CSM17	*C. pechuelii*	Namibia	-21.3760	15.0892
CSM22	*C. pechuelii*	Namibia	-21.4837	15.0390
CSM25	*C. pechuelii*	Namibia	-21.4837	15.0390
CSM28	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM29	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM31	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM33	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM34	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM35	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM36	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM37	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM39	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM4	*C. halicacabum*	South Africa	-27.6833	32.4333
CSM40	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM42	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM45	*C. grandiflorum*	South Africa	-31.4323	24.9830
CSM5	*C. corindum*	South Africa	-24.2621	29.9539
CSM6	*C. corindum*	South Africa	-24.2621	29.9539
CSM7	*C. halicacabum*	South Africa	-26.9167	32.2500
CSM8	*C. halicacabum*	South Africa	-26.8667	32.2500
CSM9	*C. halicacabum*	South Africa	na	na
CSMN4	*C. corindum*	Namibia	-20.5080	17.2472
CSMN5	*C. corindum*	Namibia	-20.5080	17.2472
CSMN6	*C. corindum*	Namibia	-20.5080	17.2472
CSMN10	*C. corindum*	Namibia	-20.5083	17.2470
CSMN11	*C. corindum*	Namibia	-20.5083	17.2470
CSMN13	*C. grandiflorum*	Namibia	-20.5084	17.2480
CSMN14	*C. corindum*	Namibia	-20.5084	17.2480
CSMN15	*C. pechuelii*	Namibia	-20.5084	17.2480
CSMN17	*C. pechuelii*	Namibia	-20.5084	17.2480
CSMN19	*C. corindum*	Namibia	-20.5095	17.2476
CSMN2	*C. grandiflorum*	Namibia	-20.5080	17.2472
CSMN20	*C. corindum*	Namibia	na	na
CSMN21	*C. corindum*	Namibia	-20.5129	17.2418
CSMN23	*C. grandiflorum*	Namibia	-20.5108	17.2425
CSMN24	*C. halicacabum*	Namibia	-20.5108	17.2425
CSMN25	*C. grandiflorum*	Namibia	-20.5108	17.2425
CSMN27	*C. halicacabum*	Namibia	-20.5038	17.2339
CSMN28	*C. corindum*	Namibia	-20.5106	17.2438
CSMN29	*C. corindum*	Namibia	-20.5106	17.2438

CSMN30	*C. corindum*	Namibia	-20.5106	17.2438
CSMN34	*C. pechuelii*	Namibia	-21.2759	15.2263
CSMN36	*C. pechuelii*	Namibia	-21.2759	15.2263
CSMN38	*C. pechuelii*	Namibia	-21.2759	15.2263
CSMN40	*C. pechuelii*	Namibia	-21.2759	15.2263
CSMN50	*C. pechuelii*	Namibia	-21.3414	15.3903
CSMN8	*C. grandiflorum*	Namibia	-20.5083	17.2470
CSMN9	*C. corindum*	Namibia	-20.5083	17.2470
FC1	*C. halicacabum*	Fiji	-18.1441	178.4428
FC2	*C. halicacabum*	Fiji	-18.1559	178.4437
JLR272	*C. grandiflorum*	Argentina	-27.5589	-57.5179
JLR278	*C. halicacabum*	Argentina	-24.2723	-64.7790
JLR280	*C. halicacabum*	Argentina	-26.4077	-54.5634
JLR281	*C. halicacabum*	Argentina	-27.4953	-56.1448
JLR282	*C. halicacabum*	Argentina	-27.4459	-57.3908
JLR284	*C. halicacabum*	Argentina	-24.0838	-64.8152
JLR286	*C. halicacabum*	Argentina	-27.5491	-57.5255
JLR288	*C. corindum*	Argentina	-23.1108	-63.7852
JLR289	*C. corindum*	Argentina	-23.1108	-63.7852
JLR290	*C. corindum*	Argentina	-23.8443	-64.0622
JLR291	*C. corindum*	Argentina	-27.2780	-55.5710
JLR292	*C. corindum*	Argentina	-27.2780	-55.5710
JLR293	*C. corindum*	Argentina	-24.2106	-64.6856
JLR294	*C. corindum*	Argentina	-24.2106	-64.6856
JLR295	*C. corindum*	Argentina	-24.2680	-61.2423
JLR296	*C. corindum*	Argentina	-24.2680	-61.2423
JLR297	*C. corindum*	Argentina	-23.2280	-63.4314
JLR298	*C. corindum*	Argentina	-23.2280	-63.4314
JLR299	*C. corindum*	Argentina	-23.2551	-63.3203
JLR300	*C. corindum*	Argentina	-23.2551	-63.3203
JLR312	*C. grandiflorum*	Argentina	-27.5829	-54.5785
JLR318	*C. grandiflorum*	Argentina	-27.3315	-55.5429
JLR321	*C. corindum*	Argentina	-24.2698	-64.8596
JLR324	*C. corindum*	Argentina	-24.0130	-61.6183
JLR326	*C. corindum*	Argentina	-23.4211	-64.1367
JLR331	*C. corindum*	Argentina	-25.9266	-64.7411
JLR335	*C. grandiflorum*	Uganda	1.1027	32.3968
JLR339	*C. halicacabum*	Australia	-13.0364	132.4397
JLR340	*C. halicacabum*	Tanzania	-5.6944	36.3223
JLRH1	*C. grandiflorum*	Hawaii, USA	21.3062	-157.80878
JLRH10	*C. grandiflorum*	Hawaii, USA	21.3401	-157.72204
JLRH11	*C. grandiflorum*	Hawaii, USA	21.3322	-157.71742
JLRH12	*C. grandiflorum*	Hawaii, USA	21.3447	-157.72290
JLRH13	*C. grandiflorum*	Hawaii, USA	21.3062	-157.80878
JLRH14	*C. halicacabum*	Hawaii, USA	21.3447	-157.81106
JLRH15	*C. grandiflorum*	Hawaii, USA	21.3053	-157.81106
JLRH2	*C. grandiflorum*	Hawaii, USA	21.3053	-157.81106
JLRH3	*C. grandiflorum*	Hawaii, USA	21.3076	-157.80903

JLRH4	*C. grandiflorum*	Hawaii, USA	21.3094	-157.80672
JLRH5	*C. grandiflorum*	Hawaii, USA	21.3094	-157.80672
JLRH6	*C. grandiflorum*	Hawaii, USA	21.3094	-157.80672
JLRH7	*C. grandiflorum*	Hawaii, USA	21.3094	-157.80672
JLRH8	*C. grandiflorum*	Hawaii, USA	21.3053	-157.81106
KZNCG4	*C. grandiflorum*	South Africa	-29.8542	30.9341
KZNCG7	*C. grandiflorum*	South Africa	-29.6083	30.3461
LR1	*C. halicacabum*	La Réunion	-21.35467	55.57233
LR3	*C. halicacabum*	La Réunion	-21.35483	35.75207
LR4	*C. halicacabum*	La Réunion	-20.91635	55.48095
May1	*C. halicacabum*	Mayotte	-12.8431	45.1383
NSWC1	*C. grandiflorum*	Australia	-33.9285	150.9633
NSWC4	*C. grandiflorum*	Australia	-33.8632	151.0114
NSWC5	*C. grandiflorum*	Australia	-33.9347	151.1326
NSWC6	*C. grandiflorum*	Australia	-33.748	151.1449
NSWC9	*C. grandiflorum*	Australia	-28.9979	153.4007
NSWC10	*C. grandiflorum*	Australia	-28.67115	153.5241
NSWC11	*C. grandiflorum*	Australia	27.0717	153.3411
NSWC12	*C. grandiflorum*	Australia	-27.4675	151.9778
NSWC13	*C. grandiflorum*	Australia	-27.4793	151.9667
NSWC16	*C. grandiflorum*	Australia	-27.5322	152.9452
NSWC17	*C. grandiflorum*	Australia	-27.522	152.409
NSWC19	*C. grandiflorum*	Australia	-27.5947	152.8449
RT10	*C. grandiflorum*	South Africa	-25.0187	31.2492
RT13	*C. grandiflorum*	South Africa	-25.0058	31.2533
RT16	*C. halicacabum*	South Africa	-25.0187	31.2492
RT19	*C. grandiflorum*	South Africa	-25.0188	31.2519
RT22	*C. halicacabum*	South Africa	-25.0188	31.2519
RT25	*C. halicacabum*	South Africa	-25.0188	31.2519
TAH1	*C. grandiflorum*	Tahiti	-17.5587	-149.5372

Chapter 3

Table S1: Sample ID, species, latitude and longitude of samples used in this study.

sample	*Species*	Latitude	Longitude
CSM15	*C. pechuelii*	-21.376	15.08917
CSM17	*C. pechuelii*	-21.376	15.08917
CSM21	*C. corindum*	-20.5073	17.2365
CSM22	*C. pechuelii*	-21.4837	15.039
CSM25	*C. pechuelii*	-21.4843	15.039
CSM5	*C. corindum*	-24.2621	29.95391
CSMN58	*C. pechuelii*	-21.3414	15.39033
CSM6	*C. corindum*	-24.2621	29.95391
CSMN20	*C. corindum*	-20.5129	17.24176
CSMN21	*C. corindum*	-20.5129	17.24176
CSMN28	*C. corindum*	-20.5106	17.24382
CSMN29	*C. corindum*	-20.5106	17.24382
CSMN3	*C. corindum*	-20.508	17.2472
CSMN30	*C. corindum*	-20.5106	17.24382
CSMN32	*C. pechuelii*	-21.2759	15.22627
CSMN34	*C. pechuelii*	-21.2759	15.22627
CSMN37	*C. pechuelii*	-21.2759	15.22627
CSMN38	*C. pechuelii*	-21.2759	15.22627
CSMN40	*C. pechuelii*	-21.2759	15.22627
CSMN41	*C. pechuelii*	-21.2759	15.22627
CSMN42	*C. pechuelii*	-21.2759	15.22627
CSMN44	*C. pechuelii*	-21.2759	15.22627
CSMN45	*C. pechuelii*	-21.2759	15.22627
CSMN46	*C. pechuelii*	-21.2759	15.22627
CSMN47	*C. pechuelii*	-21.2759	15.22627
CSMN48	*C. pechuelii*	-21.3414	15.39033
CSMN49	*C. pechuelii*	-21.3414	15.39033
CSMN5	*C. corindum*	-20.508	17.2472
CSMN50	*C. pechuelii*	-21.3414	15.39033
CSMN52	*C. pechuelii*	-21.3414	15.39033
CSMN55	*C. pechuelii*	-21.3414	15.39033
CSMN56	*C. pechuelii*	-21.3414	15.39033
CSMN58	*C. pechuelii*	-21.3414	15.39033
CSMN59	*C. pechuelii*	-21.3414	15.39033
CSMN63	*C. pechuelii*	-21.3414	15.39033
CSMN64	*C. pechuelii*	-21.8344	15.16577
CSMN67	*C. pechuelii*	-21.8344	15.16577
CSMN68	*C. pechuelii*	-21.8344	15.16577
CSMN69	*C. pechuelii*	-21.8344	15.16577
CSMN70	*C. pechuelii*	-21.8344	15.16577
CSMN72	*C. pechuelii*	-21.8344	15.16577

CSMN74	*C. pechuelii*	-21.8344	15.16577
JLR288	*C. corindum*	-23.1108	-63.7852
JLR290	*C. corindum*	-23.8443	-64.0622
JLR291	*C. corindum*	-27.278	-55.571
JLR292	*C. corindum*	-27.278	-55.571
JLR293	*C. corindum*	-24.2106	-64.6856
JLR294	*C. corindum*	-24.2106	-64.6856
JLR295	*C. corindum*	-24.268	-61.2423
JLR296	*C. corindum*	-24.268	-61.2423
JLR297	*C. corindum*	-23.228	-63.4314
JLR298	*C. corindum*	-23.228	-63.4314
JLR299	*C. corindum*	-23.2551	-63.3203
JLR300	*C. corindum*	-23.2551	-63.3203
JLR324	*C. corindum*	-24.013	-61.6183
JLR331	*C. corindum*	-25.9266	-64.7411
JLR332	*C. corindum*	-24.2723	-64.7679

Figure S1: Structure harvester results using the Evanno Method indicating the optimum number of genetic clusters (K) as 2.